Where do angry birds tweet? Income inequality and online hate in Italy

Daria Denti[a,b,] and Alessandra Faggian[a,]

[a]*Department of Social Sciences, Gran Sasso Science Institute, L'Aquila 67100, Italy, daria. denti@gssi.it, alessandra.faggian@gssi.it*
[b]*Department of Geography and Environment, London School of Economics, WC2A 2AE London, UK*

Do spatial socioeconomic features influence a digital behaviour like cyberhate? Our contribution provides an answer to this question, showing how high levels of income inequality determine high volumes of hate tweets in Italy. Our findings are robust to potential endogeneity problems of income inequality, as well as to the inclusion of confounding factors and to competing estimation strategies. Additionally, we find that education does not act as a protective factor against cyberhate in unequal places, aligning with existing evidence showing that inequality may trigger intolerance, including among educated people, threatening the perceived stability of social positions. Also, in the Italian case, the perception of economic insecurity fuels cyberhate, alongside the transmission of self-interest values along family generations. The latter finding relates to existing evidence supporting the role of persistent social norms in shaping people's attitudes.

Introduction

Online hate, also labelled as 'cyberhate', is a fast-growing phenomenon. Data show that, in the European Union, more than 75% of Internet users witnessed some sort of online hate speech on digital social platforms (Eurobarometer, 2016); in the USA, the share is about 66% (Duggan, 2017). Contrary to common belief, the data also show that the majority of cyberhate comes from people unrelated to any organised hate group (Hall, 2013). This worrying trend has pushed online hate up the research (Müller et al., 2018; Silva et al., 2016) and policy (Gagliardone et al., 2015) agenda.

Despite the growing interest in cyberhate, little is known about what really drives it, in particular about the under-investigated relationship with the geography of income inequality that is the focus of the present paper. Qualitative studies provide descriptive evidence on the association between cyberhate and several other spatial features, such as crime and voting (Bernatzky, Costello and Hawdon, 2021; Görzig, Milosevic and Staksrud, 2017), suggesting that space matters in shaping this digital behaviour (Näsi et al., 2017). Recent

quantitative evidence confirms the role of space, detailing the effect of the geography of unemployment (Anderson, Crost and Rees, 2020), human capital (Chan, 2019) and pandemic shocks (Lu and Sheng, 2020) on cyberhate. Overall, these contributions suggest that relevant risk factors for cyberhate belong to the socioeconomic context. The aim of the paper is to contribute to this strand of literature by adding evidence on income inequality, focussing on the production of hate tweets in Italy.

The effect of income inequality on cyberhate appears to be particularly salient in light of the mounting evidence showing that higher inequality relates to 'close-but-different' behaviours with regard to cyberhate. According to existing work, inequality shapes the observed patterns of political discontent (Burgoon et al., 2019; Dijkstra, Poelman and Rodríguez-Pose, 2019; Engler and Weisstanner, 2020; Martin et al., 2018; McCann, 2020; Rodríguez-Pose, 2018), and several discriminatory violent behaviours such as racial school bullying and violence against minorities (Decelles and Norton, 2016; Elgar et al., 2013; Kunst et al., 2017; Wilkinson and Pickett, 2017).

Our results support the idea that income inequality matters in fostering cyberhate, even after controlling for the spatial features that are already known to have potentially competing effects, such as crime, unemployment and immigration. Additionally, we find that the interaction between economic inequality and education is positively related to an increase in cyberhate. This counterintuitive finding parallels the existing evidence showing that educated people display more intolerant behaviours than less educated people in situations characterised by inequality (Jetten et al., 2017; Kunstman, Plant and Deska, 2016; LeBlanc, Beaton and Walker, 2015; Sharma, 2015). It is also consistent with evidence from Italy outlining non-negligible levels of bias against immigrants among educated groups as showed

by Alesina et al. (2018) in the assessment of prejudices among Italian teachers. This relates to recent findings detailing how the effect of education on hate varies by country (Finseraas, Skorge and Strøm, 2018; Weber, 2020), with evidence showing a protective effect of education in some places, but not in others. Finally, we find that economic insecurity and social norms promoting self-interest act as risk factors for cyberhate, whereas neither crime nor immigrants have meaningful association with it.

Overall, the economic dimension is relevant in shaping cyberhate, in line with many hate-studies (that is, Gerstenfeld, 2017; Green, Mcfalls and Smith, 2001; Stephan and Stephan, 2000). The evidence provided by our paper is important for policy, since it identifies which factors should be addressed to counter cyberhate and shows that some of these could be tackled and/or alleviated without imposing a regulation interfering with the fundamental right to freedom of speech (McGonagle, 2013).

We focus on the case of Italy for both practical and theoretical reasons. First, Italy is one of the most unequal countries among the Western European countries (Eurostat, 2019) and it experiences volumes of online hate far higher than the European average. Nearly 80% of Italian internet users witnessed some sort of online hate speech (SWG, 2017). Second, we have very detailed data on cyberhate in Italy, consisting of more than 75,000 geo-referenced hate tweets collected in 2017 (Musto et al., 2016), which we merged with administrative data on the 611 Italian Local Labour Market Areas (LLMAs). Estimating a two-part model (2PM) and a Control Function with Two-Stage Residual Inclusion (CF-TSRI), we find that the volume of cyberhate is determined by income inequality, even when we control for other potential confounding factors and for endogeneity issues. The results appear robust to alternative model specifications (single-stage ordinary least squares [OLS] and sample

selection model [SSM]), as well as to a series of other robustness checks.

Our contribution builds on previous research along several dimensions. First, our results contribute to the build-up of the empirical evidence targeting the links between places and cyberhate, as called for by hate scholars (Gagliardone et al., 2015). Our findings provide quantitative support for the role of the socio-economic dimension, adding income inequality as relevant determinant for cyberhate to the existing evidence on unemployment, shocks and education (Anderson, Crost and Rees, 2020; Chan, 2019; Lu and Sheng, 2020). Second, our results relate to the literature detailing the strong role of places in shaping resentment (Abreu and Öner, 2020; Billing, McCann and Ortega-Argilés, 2019), adding evidence on the specific behaviour of cyberhate. Third, this investigation refers to the thriving research on the role of inequality in influencing resentful conducts (Burgoon et al., 2019; Côté, House and Willer, 2015; Engler and Weisstanner, 2020; Wilkinson and Pickett, 2017), by adding evidence on cyberhate. Fourth, our findings also relate to novel evidence suggesting that the effects of education in promoting tolerance entail a relevant country-specific dimension (Alesina et al., 2018; Finseraas, Skorge and Strøm, 2018; Lancee and Sarrasin, 2015), which can be influenced by the level of inequality (Piff and Moskowitz, 2017; Wodtke, 2016). Fifth, given the focus on the role of spatial geographies on online hate, the paper contributes to the debate on the role of places in the ICT-driven world, showing that places actually matter in shaping online behaviours.

The remainder of the paper is organised as follows. Section Drivers of online hate speech describes cyberhate, its potential association with economic inequality and with other spatial features. The Results section introduces the data and the empirical strategy. Results are presented and discussed in the Discussion section. Finally, the fifth section concludes.

Drivers of online hate speech

Online hate speech: definition, characteristics and spatial risk factors

Online hate speech is defined as: 'words or symbols diffused through the Internet, that are derogatory and/or intimidating on the basis of race, religion, sexual orientation, and so on' (McGonagle, 2013). Currently, it is one of the challenges posed by the extensive use of social media (European Commission, 2018; OSCE-ODIHR, 2010). The impact of cyberhate on victims can be devastating and durable, due to hyperlinking, online searchability and content shared by users (Sunstein, 2017). Online hate is extremely pervasive and it becomes public and available to a potentially global audience without either mediation or cost. In 2018, Facebook removed around 7.9 million pieces of content related to hate speech worldwide (Facebook, 2019), and YouTube cancelled more than 160 channels per day globally (Youtube, 2019); Twitter has deleted nearly 2.5 million tweets for hateful contents in 2019 (Twitter, 2019). The relevance of cyberhate is confirmed also by the ongoing trend towards law enforcement to counter it (Assimakopoulos et al., 2017). Online hate occupies a prominent position also in numerous research fields, from computer science and criminology to economics and psychology (Gagliardone et al., 2015; Müller et al., 2018; Silva et al., 2016), and it is classified as a stand-alone resentful oppressive behaviour, structurally different from offline hate (ElSherief et al., 2018; Hine et al., 2016; Müller et al., 2018).

Evidence shows that real-world social and moral norms which are effective in moderating real-world hate have a remarkably weak grip in countering online hate (Lowry et al., 2016). This is due to several reasons. First, cyberhate is, by definition, created online, where people aggregate in homogeneous clusters (Himelboim et al., 2013). By creating these virtual in-groups, users reinforce their extant social group identity and they create '*echo chambers*', which are

defined as closed systems where stereotypes and prejudices are amplified and reinforced (Sunstein, 2017). Overall, the distance between ethnic/cultural groups increases. Second, there is a widespread uncritical acceptance of information found on the internet, which further reinforces prejudice and social identity (Hall, 2013). Third, the online hatemongers experience an '*online disinhibition effect*' (Suler, 2004) that lessens their sense of accountability and moderation (Citron et al., 2011) and lowers the effect of real-world social barriers in countering aggressive and radical behaviours (ElSherief et al., 2018; Sunstein, 2017). The '*online disinhibition effect*' results from a perceived sense of anonymity experienced by many online users (Perry et al., 2009).

Although more shielded from real-world social norms, cyberhate does not happen in a spatial vacuum (Castells, 2001; Hall, 2013), since hatemongers are grounded in a spatial milieu. Cross-country descriptive evidence outlines place-specific heterogeneity in cyberhate, also when controlling for individual characteristics (Näsi et al., 2017). Qualitative evidence shows that spatial features—such as crime, unemployment, social capital and hate events—relate to online hate (Bernatzky, Costello and Hawdon, 2021; Costello and Hawdon, 2018; Görzig, Milosevic and Staksrud, 2017; Kaakinen et al., 2018; Kowalski, Limber and McCord, 2019). Novel quantitative evidence provides further support on the influence of local features on cyberhate. Anderson et al. (2020) show that US counties with higher levels of unemployment display higher shares of cyberhate. Chan (2019) finds a correlation between the local endowment of educated people and cyberhate, again in the USA (but provide no support for a causal link). Lu and Sheng (2020) identify a causal link between the local spread of Covid-19 and cyberhate.

This growing evidence gives room to further analyse the relationship between cyberhate and places. We contribute to this investigation providing evidence on the role of income inequality, given the broad empirical literature bears out a strong association between income inequality and other anti-social attitudes, but there is a lack of quantitative analysis specifically targeting cyberhate.

Income inequality and online hate

Evidence shows that income inequality relates to resentment, racial school bullying and violence against minorities.

Income inequality strongly predicts voting for parties proposing anti-immigrant/anti-global platforms (Burgoon et al., 2019; Engler and Weisstanner, 2020; Gest, Reny and Mayer, 2018). Discontent expressed in the ballot box also relates to territorial socioeconomic inequality, as detailed in the growing bulk of literature on the 'places that don't matter' (Rodríguez-Pose, 2018, 2020). Observed patterns of resentment expressed in the ballot box are associated with the geographical imbalance between places that thrive from globalisation and places that feel left-behind (Billing, McCann and Ortega-Argilés, 2019; Dijkstra, Poelman and Rodríguez-Pose, 2019; Iammarino, Rodriguez-Pose and Storper, 2019; Martin et al., 2018; McCann, 2020).

Cross-country evidence shows a strong association between inequality and several discriminatory behaviours: racial school bullying (Elgar et al., 2013), lack of empathy and solidarity towards minorities, scapegoating and homophobia (Andersen and Fetner, 2008; Côté, House and Willer, 2015; Layte and Whelan, 2014; Wilkinson and Pickett, 2017). Data from the US details how higher levels of inequality relate to more hate, triggering racism, sexism, opposition to social welfare and violent acts against minorities (Kunst et al., 2017).

We build on these findings with our investigation on the effect of income inequality on cyberhate in Italy. Specifically, we focus our analysis on disposable income inequality. Our starting point is the assessment of the measure of association between disposable income inequality and cyberhate, considering also

other confounding factors that are related to cyberhate according to the existing literature. A detailed description of these competing spatial factors is provided in the next subsection.

We also explore whether the effect of disposable income inequality on cyberhate changes depending on the local share of educated people, given the mixed evidence on this. Experimental evidence from Australia, Canada, USA and India outline that more educated people display more intolerant behaviours when economic inequality is higher (Decelles and Norton, 2016; Kunstman, Plant and Deska, 2016; LeBlanc, Beaton and Walker, 2015; Sharma, 2015). Scholars explain these findings through the 'fear of falling' effect triggered by income inequality (Jetten et al., 2017). In other words, income inequality makes higher-status people more concerned about losing their privileged position, pushing them to develop legitimising frames for scapegoating disempowered groups (Lick, Alter and Freeman, 2018; Piff and Moskowitz, 2017; Wodtke, 2016). The evidence supporting the 'fear of falling' effect in some countries is in sharp contrast to the findings in other countries, where higher educated people are more tolerant towards minorities (Cavaille and Marshall, 2019; Lancee and Sarrasin, 2015) in unequal places as well (Korndörfer, Egloff and Schmukle, 2015). This evidence supports education as a way for better framing the threats faced by society and internalising equality values (Cavaille and Marshall, 2019). Given these contrasting results, which effect prevails in a given country appears to be a matter for empirical investigation.

Recent works scrutinising the influence of education on hate provides support for country-specific effects. Data from Switzerland shows that higher educated individuals become more likely to have discriminatory attitudes when they enter the labour market (Lancee and Sarrasin, 2015) and evidence from Italy indicates that an educated group, such as teachers, display strong negative stereotypes towards immigrant students to the point of influencing marking (Alesina et al., 2018). Empirical evidence shows little support for high levels of education in decreasing hate also in Germany (Weber, 2020), Norway (Finseraas, Skorge and Strøm, 2018), UK, Sweden (Cavaille and Marshall, 2019) and the USA (Wodtke, 2016). Data from experimental games targeting USA and UK indicates that people with high educational attainment display a stronger individualistic behaviour (Manstead, 2018), and reduced empathy and trust (Kraus, Côté and Keltner, 2010), after controlling for personality traits. At the same time, there is broad and established evidence on education as having an effective impact in countering hate at the European level (*i.a.* d'Hombres and Nunziata, 2016; Rooduijn, 2018), and in several countries including France and the Netherlands (Cavaille and Marshall, 2019). Qualitative evidence suggests that education counters the perception of economic threats, but it is less effective in countering the perception of a sociocultural threat (Schmuck and Matthes, 2015), relating to the fear of losing social status.

The observed cross-country heterogeneity on the effects of education and of its interaction with income inequality on attitudes towards minorities suggests that the effect which prevails in a given country is a matter for empirical investigation (Cavaille and Marshall, 2019; Finseraas, Skorge and Strøm, 2018), which we pursue, in the present paper, for Italy.

Finally, we also account for the potential reverse causality in the measure of association between income inequality and cyberhate, referring to the literature addressing income inequality and public bads (Enamorado et al., 2016). The public bad nature of cyberhate may stimulate selective outmigration of affluent people from places characterised by high level of intolerance and this potential bias might affects our findings. To account for this, we adopt an instrumental variable approach, exploiting a Bartik-type exogenous regressor following established contributions on endogeneity bias and income inequality (Baum-Snow and

Ferreira, 2015; Boustan, Ferreira, Winkler and Eric M. Zolt, 2013; Enamorado et al., 2016).

Other spatial features related to cyberhate

Existing studies detail several spatial features capable of influencing cyberhate production besides inequality, which we describe below. We will include them in our empirical investigation as control variables to assess the robustness of our results and to contribute to the information base on the role of spatial features on cyberhate, with evidence from Italy.

Family is an interesting element to consider given its focal imprinting on the spatial geographies of the transmission of equality/non-equality value (Alesina et al., 2021; Bertocchi a et al., 2019; Duranton et al., 2009) and the acknowledged role of non-equality values in reducing solidarity (Wilkinson and Pickett, 2017). In accounting for the geography of value transmission through family, we rely on the established classification of family types by Todds (1990), who identified an organising principle for the classification of family types along the equality/inequality dimension based on the relationship between siblings in the family, as shaped by what happens to family property after the death of the parents. Equality is said to be strongest where family property is divided most evenly between siblings, whereas it is weakest when one particular child (often the eldest) is favoured at the expense of the others. Areas in which equal familial systems are operating are identified, therefore, by inheritance laws and practices. The classification of family types identifies two egalitarian family types—*communitarian* and *egalitarian nuclear*—and two non-egalitarian family types—*stem* and *incomplete stem* (Todds, 1990). Non-egalitarian family types are identified by the self-interest dimension enforced through individualistic standards. All four types of family are present in the Italian context (Duranton et al., 2009). Notably, the classification of family

types along this dimesion has already been acknowledged as capable of influencing relevant socioeconomic outcomes (Duranton et al., 2009). The same individualistic standards channelled by the non-egalitarian family types are also recognised within social psychology as a booster for social anxiety and violence, since they increase the relevance of preserving the existing social status from potential threats (Wilkinson et al., 2017). Thus, we bridge these two strands of literature by assessing whether self-interest values transmitted through non-egalitarian family types influences online hate.

Perceived competition for scarce resources, such as jobs and welfare, can act as triggers for hateful behaviour (Green et al., 2001; Stephan et al., 2000). Experimental evidence shows that perceived scarcity brought about by economic hardship affects people's representations of minorities fostering discrimination (Krosch and Amodio, 2014; Krosch, Tyler and Amodio, 2017), and quantitative evidence identifies unemployment and the perception of job precariousness as determinants of online hate in the USA (Anderson, Crost and Rees, 2020). The perception of job precariousness is particulary interesting in our analysis, given that data from Italy acknowledges that its role in influencing people's discontent and behaviours is more relevant than fluctuations in actual employment figures (Boeri and Brandolini, 2005; Modena, Rondinelli and Sabatini, 2014; OECD, 2018).

Social capital can influence cyberhate, since lower levels of trust promote disconnectedness among different social groups (Gerstenfeld, 2017). Conversely, high levels of trust strengthen the adherence to social norms and promote pro-social solidaristic behaviours (Andriani and Sabatini, 2015). Collaboration favours openess towards diverse groups. Cross-country evidence supports a negative association between offline trust and cyberhate (Kaakinen et al., 2018), whereas the descriptive findings on the association between cyberhate and collaboration are mixed (Hawdon et al.,

2020; Kaakinen et al., 2018), probably due to the nature of collaboration (among people belonging to the same social group or among people belonging to different social groups).

We also consider crime, since it is one of the main sources for social tension and distress (Dustmann and Fasani, 2016; Pinotti, 2015;) and qualitative evidence suggests an association with cyberhate (Görzig, Milosevic and Staksrud, 2017). Similarly, foreign population and refugees might work as a risk factor for cyberhate, since immigrants may be perceived as a threat to the sociocultural identity of the locals (Bansak, Hainmueller and Hangartner, 2016; Hainmueller and Hopkins, 2014) and immigrants are one of the main targets of cyberhate (OSCE-ODIHR, 2019). Votes for anti-minorities political platforms are a relevant proxy for the social unrest determined by the geography of winners and losers from globalisation (Martin et al., 2018; Rodríguez-Pose, 2020). Novel qualitative evidence on cyberhate indicates that this form of unrest may fuel online hate narratives occurring in the same areas (Bernatzky, Costello and Hawdon, 2021). By the same argument, real-world hate crimes constitute another potential risk factor, by contributing to legitimising online anti-minorities behaviours, as also suggested by recent qualitative evidence (Costello and Hawdon, 2018).

Data

We measure cyberhate using the corpus of Twitter geo-referenced data extracted through a system of algorithms designed by Musto et al. (2016) and used to design the *Italian Hate Map*,[1] in turn inspired by the *Humboldt University Hate Map* targeting the USA.[2] The database contains more than 75,000 tweets generated in Italy in 2017 and identified through data extraction algorithms targeting semantic processing, sentiment analysis and content classification. By hate tweet we refer to sentences posted on Twitter containing at least one derogatory term used in a violent and/or derogatory way against people on the basis of ethnicity, sexual

orientation, gender or disability. The corpus of hate tweets has been designed through several steps, consistent with the computer-science literature on hate tweet detection (i.a. Burnap and Williams, 2014; Himelboim, Mccreery and Smith, 2013). First, a set of 47 sensible terms is identified for the following intolerance dimensions: homophobia, racism, violence, disability, anti-Semitism, gender. Second, an algorithm extracting the Italian tweets containing at least one of the sensible terms is launched, extracting tweets for 10 months. Third, the extracted tweets are analysed to remove non-intolerant tweets, that is, tweets that contain a sensible term but used in a non-hateful message. This third part discards those tweets that are characterised by neutral and positive sentiment, leaving only tweets with at least one sensible term and a negative sentiment against minorities. Fourth, of the remaining tweets those which are geo-tagged are retained (Musto et al., 2015). This multi-stage process is necessary to discard false positives.

Although capturing only a part of cyberhate, Twitter is a valuable source, being widely used to propagate hate, as well as being characterised for having public content that can be retrieved and analysed (ElSherief et al., 2018; Himelboim et al., 2013). We aggregate the hate tweets at Local Labour Market Area (LLMA) and normalise them by the total tweets generated.[3] Figure 1 portrays the resulting geography. The 611 Italian LLMAs divide Italy in functional areas based on commuting, hence containing the bulk of the labour force living and working there. They are particularly suitable to our study because they alleviate the unfeasibility in detecting whether the tweets are posted during working time, commuting time or leisure time. Hate-related tweet extraction does not include the general flow of tweets, which we measure using the Cheng et al. (2011) Twitter corpus, up to now the largest and finest-grained geo-tagged Twitter database available.

We measure disposable income inequality across the 611 LLMAs via a Gini index (Acciari and Mocetti, 2013).
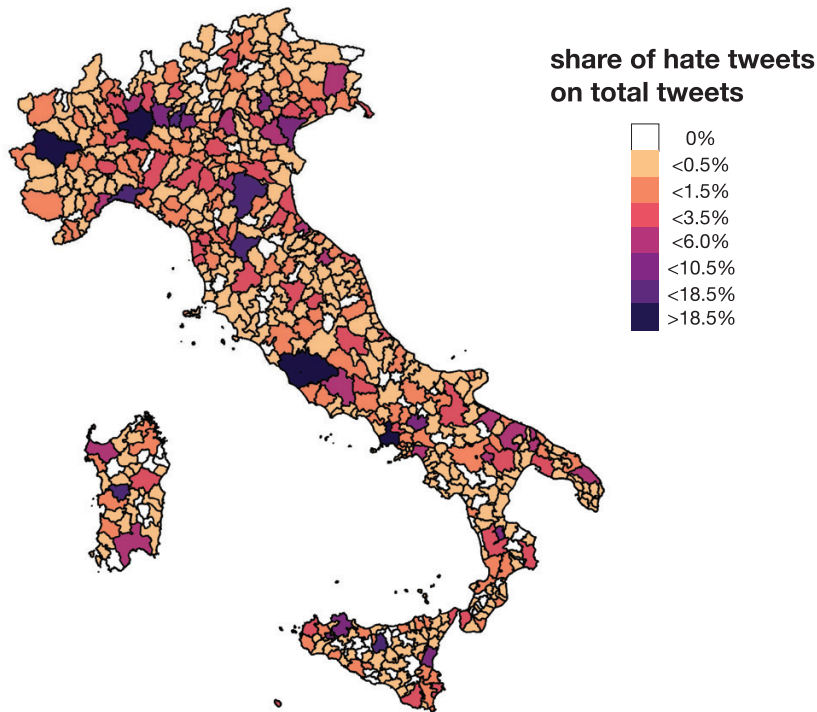
**share of hate tweets
on total tweets**

☐	0%
	<0.5%
	<1.5%
	<3.5%
	<6.0%
	<10.5%
	<18.5%
	>18.5%

Figure 1. *The map of hate tweets across the 611 Italian Local Labour Market Areas (LLMAs).*

To account for concerns on the endogeneity of disposable income inequality, later we introduce an instrumental variable to predict the income distribution of LLMA *i* at time *t* using 16-year lagged information on the local disposable income distribution and national growth rate for each income bin. This Bartik-type instrument for the Gini index of income inequality has been introduced by Boustan et al. (2013) and further applied in the literature addressing the endogeneity bias of income inequality (Baum-Snow and Ferreira, 2015; Enamorado et al., 2016).

To measure the local transmission of either egalitarian or non-egalitarian values, we map the geography of egalitarian and non-egalitarian families following Duranton et al. (2009) on the Italian LLMAs. Within the economic dimension, we consider proxy economic insecurity with a measure for job precariousness given the salience of this features on several behavioural outcomes for Italians. We consider the share of workers feeling insecure about being able to keep their job (ISTAT, 2016).

We account for the other potentially confounding factors acknowledged by the literature. For the demographic dimension, we consider the share of resident migrants, the geography of refugee hosting centres, non-hate and hate crime. On the social dimension, we measure trust through the voting turnout at the 2014 European Parliament elections, and collaboration through the number of non-for-profit local units and through vaccine coverage that proxies for the attitude of putting collective needs before individual needs (WHO, 2018; Wolfe, 2002; Kennedy, 2019). Political preferences for anti-immigrant parties are captured by

the share of votes for right-wing parties in the 2014 European Parliament elections. The empirical investigation also controls for population size, population density, the share of disrupted families and central vs marginalised areas. Summary statistics for the considered variables are included in Table 1,[4] the data description and the data sources are in Supplementary Table A.1 and A.2.

Empirical modelling

Our baseline model specification focuses on the correlation between income inequality and the share of geotagged online hate tweets generated in 2017 across the 611 Italian LLMAs. The database is cross-sectional and the dependent variable, given by the share of geo-tagged hate tweets on total geo-tagged tweets produced in each LLMA, is

Table 1. *Descriptive statistics*

Variable	Obs	Mean	SD	Min	Max
Geography of hate tweets					
Onlinehate tweets	1	0.000663	0.00185	0	0.0261
Volume of total tweets	1	0.858	0.350	0	1
Economic characteristics					
Human capital	611	10.28096	2.737488	3.578903	21.07972
Gini index	611	41.135	2.866571	33.47302	54.87072
Company failures	611	12.43	5.680	1	30.10
Overdue payments	611	23.45	11.47	4.400	48
Perception of job security	611	5.473344	0.7065942	3.571887	7.629317
Sociocultural features					
Foreign resident population	611	5.799721	3.777264	0.2714932	16.39703
Split population	611	3.770977	1.509572	0.7702183	8.319397
Aging index	611	16.69316	4.646057	6.93617	34.9384
Population density	611	2.056063	2.915807	0.1041765	31.07485
Population	611	0.09727	0.25829	0.314	3.6851
Geographic characteristics					
Distance from closest refugees' hosting facility in 2010	611	3.6711	2.8916	0.0777543	19.23096
Hub in the LLMA	611	0.3698854	0.483169	0	1
Central Area	611	0.3927987	0.4887728	0	1
Social structures					
Social networks	611	5.528334	2.32694	0.9486006	14.95564
Trust	611	56.41075	12.89531	25.47039	85.69295
Family types	611	1.92144	1.275384	1	4
Deviant behaviours					
Crime	611	19.93512	7.503294	7.467119	55.05622
Offline hate	611	0.270	0.444	0	1
Gambling availability	611	2.242603	0.6932691	0. 340329	5.066981
Measles vaccine coverage	611	93.96	3.659	82	100
Political preferences					
Right parties votes	611	27.23696	8.619548	5.359276	55.28928
Left parties votes	611	44.17219	9.47298	9.82249	73.67038
5-Star votes	611	20.27569	5.679561	1.990223	38.42243
Center parties votes	611	6.647648	5.036799	0.2094972	48.85226
Exclusion restrictions (SSM)					
No internet age index	611	3.716	1.134	1.372	8.266
No internet age index robustness check	611	2.528	0.828	0.865	5.961

continuous with a non-negligible share of zero (as shown in Supplementary Figure A.1). The share of hate tweets displays a substantial skewness, with a long thin right tail (also see Supplementary Figure A.1). We assume that the zeros in the outcome variable are not driven by any selection bias; hence, there are only true zero observations.

Given this assumption, our preferred modelling strategy is a 2PM. First, we estimate the occurrence of hate tweets by means of a probit model for the first part *Pr(onlinehate > 0)*; second, we estimate a log-normal model for hate tweets given that some hate tweets are present, *E(ln onlinehate |onlinehate>0),* where log-normality accounts for correcting the right-skewness of the dependent variable once the condition *onlinehate >0* is applied (Cameron et al., 2010). Formally, let ln *onlinehate* be the log of the share of hate tweets on total tweets in each LLMA in 2017 and let *z* be the binary indicator of positive online hate events such that *z = 1* if *onlinehate > 0* and *z = 0* if *onlinehate = 0*. Then, for *onlinehate > 0*, f (ln *onlinehate*| $z = 1$) is the conditional density of ln *onlinehate*. Hence, the 2PM can be summarised as follows:

f (ln *onlinehate* | X) =

$$\begin{cases} Pr(z = 0|X), \ if \ onlinehate = 0 \\ \Pr(z = 1 \mid X) \ f(\ln onlinehate|z = 1, \ X), \ if \ onlinehate > 0 \end{cases}$$
(1)

$$z_i = \alpha + \beta_1 GINI_i + \beta_2 EDU_i + \beta_3 GINI_i \times EDU_i \\ + \beta_4 FAM_i + \beta_5 INS_i + \beta_6 \ \Omega_i + \varphi_i + \varepsilon_i$$
(2)

$$lnonlinehate_i = \delta + \gamma_1 GINI_i + \gamma_2 EDU_i + \gamma_3 GINI_i \times \\ EDU_i + \gamma_4 FAM_i + \gamma_5 INS_i + \\ \gamma_6 \ \Omega_i + \mu_i + u_i$$
(3)

where equation (1) summarises both stages of the 2PM, equation (2) is the stage 1 probit model and equation (3) the stage 2 log-linear regression. In equations (2) and (3), $GINI_i$ is the Gini coefficient measuring disposable income

inequality in LLMA *i*, EDU_i is the share of educated people in the same LLMA, FAM_i is the family type characterising the LLMA and INS_i is the perception of economic insecurity. In both stages we consider the interaction between income inequality and the share of educated people to measure whether the latter serves as moderator for the effect of income inequality. Ω_i contains other potential control variables. Both stages include a regional fixed effect, respectively φ_i in the probit and μ_i in the log-linear regression. ε_i and u_i are the error terms. Province, instead of regional, fixed effects are included in the robustness checks. The independent processes of the 2PM allow for the flexibility of having different regressors in the two equations. We exploit this feature by testing the same broad set of controls in Ω_i in both stages, which we assess through postestimation diagnostics to identify the relevant ones in each stage.

However, the 2PM model does suffer from possible endogeneity issues, due to reverse causality, omitted variables and/or measurement errors. To solve this endogeneity problem, we adopt a CF-TSRI approach (Stock, 2001; Terza, 2017; Wooldridge, 2015).

The CF-TSRI approach estimates a reduced-form equation where the potentially endogenous variable, that is, income inequality, is regressed against an extra regressor and Ω_i. The extra-regressor is built, for the reduced-form equation, instrumenting the Gini index with the synthetic inequality measure expressed through the Bartik-type regressor according to the literature (Boustan, Ferreira, Winkler and Eric M Zolt, 2013; Enamorado et al., 2016). In practical terms, we start with the initial (2001) average household income by local quintile and LLMAs. We then estimate to which national percentile of the income distribution each local income quintile belongs to in the initial year. Then, we allow the income of each local quantile to grow over time as the income of its corresponding national percentile.[5]

We argue that the instrument satisfies the exclusion restriction since the 2001 local income distribution is the only source for its cross-sectional variation and it is more than 16-year lagged with respect to the year for which we have data on hate tweets, 6-year lagged with respect to the first-ever-sent Tweet and 3-year lagged from the first Facebook profile. It is also 8-year lagged with respect to hate events turning non-sporadic in the country (Lunaria, 2019). Therefore, the instrument appears to be capable of mitigating concerns about anticipation effects on future streams of hate tweets and overall tweets as well as it can alleviate concern about sorting and migration of the population due to preferences with respect to hate. The reduced-form residuals are then plugged into the structural equation together with the endogenous explanatory variable and Ω_i.

Results

Baseline 2PM model

Table 2 summarises the findings from the baseline 2PM specification. Among family types, the egalitarian nuclear family is used as base category and therefore not included. Columns 1–2 report the estimates of stage 1 of the 2PM, where we assess the risk factors associated with the occurrence of online hate at local level, with and without the inclusion of the interaction between income inequality and education among regressors (for the detailed estimation results, see columns 1–2 in Supplementary Table A.3). The results show that neither income inequality nor its interaction with educated people, are associated with the occurrence of cyberhate. Whereas educated people and the local perception of job insecurity are risk factors for the occurrence of online hate. The increase of 1 percentage point in the perception of job insecurity (that is, keeping the current job for the next 6 months) is related to an increase greater than 0.55 in the occurrence of cyberhate. This result holds when we consider data on the perception of job insecurity at a different point in time (see column 2 in Supplementary Table A.3). The share of educated people is positively associated to the occurrence of cyberhate. We have tested several interactions between educated people and other variables, but none are significant. Similarly, we tested the interaction between inequality and the perception of job insecurity and find it non-significant. As for family types, the prevalence of non-egalitarian family types is associated with an increase in the occurrence of cyberhate, as shown by the positive coefficients of both *stem* and *incomplete stem* families. The presence of the egalitarian family type is instead a protective factor for the occurrence of online hate (negative coefficient of the *communitarian* family type). The local shares for right-wing parties display a small and positive association to the occurrence of cyberhate. Finally, neither crime nor immigrants are significant. These findings hold also with a wide array of robustness checks which we detail later on.

Columns 3–6 report the findings from stage 2, where the outcome is the 'intensity' (and not simply the occurrence) of hate tweets (Supplementary Table A.8). Estimates in columns 3, 5 and 6 show that high levels of income inequality are associated with high volumes of hate tweets, including when control variables are included. In column 4, we present the estimates when the Gini index is not included among regressors to show that an increase in the share of educated people is positively associated with an increase in cyberhate intensity. Column 5 shows the estimates when both educated people and the Gini index are included. Results from columns 3–6 show that the direct effects of both income inequality and education are positive and significant and that a higher level of inequality (for example, Gini index 0.01 higher) is related to a 5-p.p. change in hate tweets. Column 6 reports the estimates of the interaction between the Gini index and the share of educated people.

Table 2. *2PM results*

	2PM stage 1: probit model (y = online hate: Y/N)		2PM stage 2: log-linear model (y = share of online hate in logs)				
	(1)	(2)		(3)	(4)	(5)	(6)
	coef	coef		coef	coef	coef	coef
Job Insecurity	0.559*** (0.128)	0.556*** (0.135)					
Educated People	0.159*** (0.051)	0.161*** (0.049)	Educated People		0.495*** (0.036)	0.273*** (0.039)	0.260*** (0.041)
Gini index	0.346 (0.263)	0.119 (1.307)	Gini index	15.943*** (3.208)		5.002** (1.983)	5.784** (2.541)
Educated People *Gini index		0.0281 (0.165)	Educated People *Gini index				1.570** (0.587)
Incomplete Stem Family (*Non-Egalitarian*)	5.466*** (0.694)	5.422*** (0.715)	Incomplete Stem Family (*Non-Egalitarian*)	0.886*** (0.119)	0.227** (0.074)	0.306 (0.083)	0.361 (0.756)
Stem Family (*Non-Egalitarian*)	5.142*** (1.027)	4.834*** (1.103)	Stem Family (*Non-Egalitarian*)	1.405*** (0.093)	0.442 (0.320)	0.596 (0.161)	0.653 (0.667)
Communitarian Family (*Egalitarian*)	−0.793* (0.422)	−0.776* (0.436)	Communitarian Family (*Egalitarian*)	0.294 (0.303)	0.478 (0.387)	0.319 (0.296)	0.823 (0.934)
Right parties' votes	0.0343** (0.021)	0.0348** (0.0165)					
Crime rate	0.189 (0.414)	0.192 (0.418)	Crime rate	0.524 (0.307)	0.619 (0.429)	0.599 (0.315)	0.555 (0.339)
Foreign population	−0.0560 (0.0425)	−0.0568 (0.0421)	Foreign population	−0.007 (0.032)	−0.008 (0.028)	−0.0033 (0.028)	−0.0081 (0.029)
Observations	611	611	Observations	524	524	524	524
Regional FE	*YES*	*YES*	Regional FE	*YES*	*YES*	*YES*	*YES*
			R squared	0.298	0.306	0.396	0.440
Control: population size, vaccine coverage, offline hate			Control: offline hate, marginalised area				

Standard errors in parentheses are clustered at regional level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$; results hold removing crime rate and foreign population. Other potential confounders have been tested resulting never as significant predictors, including ageing index, population density, distance from closest refugees hosting centre, trust. See section 5.4 and Supplementary Table A4 for a detailed description of the procedure to exclude non-robust confounders.

The estimates support the significance of an indirect effect of income inequality channelled by the share of educated people in the LLMA, as shown by the positive and significant coefficient of the interaction term.[6] Figure 2a describes the effect of the interaction between income inequality and educated people on the intensity of hate tweets summarising the predictive margins. Interacting with higher inequality, human capital endowment relates to higher intensity of cyberhate as portrayed by the upward slopes. These results are robust to several robustness checks. Overall, we find that income inequality acts as a significant risk factor for the intensity of online hate.

CF-TSRI model for the endogeneity of income inequality

The 2PM estimates show correlation, but not necessarily causation, between inequality and cyberhate. In fact, they do not account for possible endogeneity problems. To account for
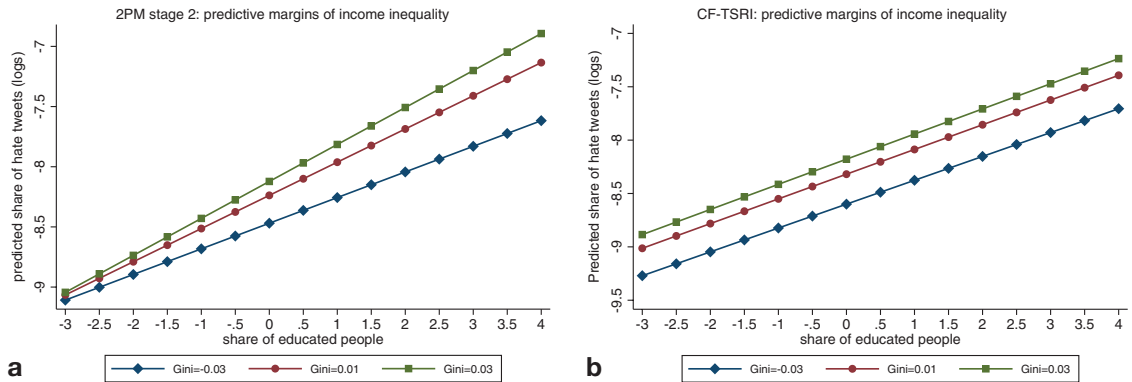
Figure 2. *Income inequality and cyberhate: the effect of human capital. (a) Marginal effects of income inequality: 2PM model (different levels of human capital). (b) Marginal effects of income inequality: CF-TSRI model (different levels of human capital).*

these, following a well-established literature on non-linear models, we use a CF-TSRI approach. Our model has an interaction term involving inequality, which might be partially correlated with inequality itself. Following Wooldridge (2010), we deal with this introducing two reduced-form equations in the CF-TSRI estimation. In both reduced-form equations, our chosen instrument is given by the Bartik-type instrument, that predicts the actual Gini index as a weighted average of national patterns of income growth (the 'shift' in the literature on Bartik-type instruments) using as weights the *i*th LLMA's income distribution in 2001 (the 'shares' in the literature on Bartik-type instruments). The first reduced-form equation regresses $GINI_i$ on the Bartik-type instrument, educated people, the interaction between the instrument and educated people and control variables. The resulting residuals are saved. Then, we regress the interaction term $GINI_i \times EDU_i$ on the Bartik-type instrument, educated people, the interaction between the instrument and educated people and control variables. As before, we save the residuals. Finally, the residuals from both reduced-form equations are included in the estimation of the structural-form equation, which is given by equation (3). Table 3 shows the results of the CF-TSRI.

Column 1 reports the CF-TSRI estimates of the structural equation with non-meaningful covariates (crimes and foreign population) not included in the regression. All our main findings are also confirmed after controlling for the potential endogeneity of income inequality (Supplementary Table A.11). The estimated coefficient for the Gini index is still positive and highly significant. Hence, high level of income inequality in LLMAs determines high volume of online hate tweets. In fact, after accounting for endogeneity the magnitude of the coefficient is even higher: a 1-p.p. increase in the Gini index implies a percentage change of more than 10% in hate tweets. The significance of the interaction between educated people and income inequality is also confirmed, as shown by the positive value for estimated coefficient for the interaction term. Therefore, income inequality acts as a determinant for the volume of hate tweets through both a direct and an indirect effect, where the latter is channelled through the local endowment of human capital and summarised in Figure 2b. Figure 2b shows the marginal effect of an increase in the Gini index of income inequality on the share of hate tweets for different shares of local human capital endowments, confirming that higher levels of human capital endowment

Table 3. *Estimates for the CF-TSRI approach applied to the stage 2 of the 2PM.*

CF-TSRI 2PM stage 2: Structural-form equation (y = share of online hate)		CF-TSRI 2PM stage 2: Reduced-form equations	
	(1)		(2)
	coef		coef
		a: direct effect (y = Gini index)	
Educated people	0.319***	Educated people	0.0437***
	(0.0403)		(0.0046)
Gini index	10.28***	Synthetic Gini index	1.053***
	(2.577)		(0.261)
Educated people × Gini index	2.982**	Educated people × Synthetic Gini index	0.197
	(0.940)		(0.117)
Incomplete stem family	2.301**	Controls	*YES*
(*Non-Egalitarian*)	(0.567)	Observations	524
Stem family (*Non-Egalitarian*)	3.344***	Regional FE	*YES*
	(0.903)	*R* squared	0.490
Communitarian family	2.987**	*F*-stat ex instrument	16.21
(*Egalitarian*)	(1.581)		
		b: interaction (y = educated people × Gini index)	
Reduced eq residuals (direct effect)	−6.421***	Educated people	0.076*
	(1.892)		(0.039)
		Synthetic Gini index	2.550
Reduced eq residuals (interaction)	−1.179***		(2.318)
	(0.188)	Educated people × Synthetic Gini index	0.258**
			(0.075)
Controls	*YES*	Controls	*YES*
Observations	524	Observations	524
Regional FE	*YES*	Regional FE	*YES*
R squared	0.458	*R* squared	0.126
		F-stat ex instrument	10.27
Control: offline hate, marginalised area		Control: offline hate, marginalised area, family types	

Standard errors in parentheses are clustered at regional level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

increase the intensity of cyberhate for any level of inequality. Column 2 reports the estimates from the reduced-form equations. The top section of column 2 describes the results of the reduced-form equation with Gini index as dependent variable. The bottom section presents the results of the other reduced-form equation with the interaction between the Gini index and educated people as the dependent variable. Estimates from column 2 show that the synthetic Gini index does predict the actual income inequality consistently with existing evidence on the positive growth of the Gini index for income inequality in Italy in our considered time span (Acciari and Mocetti, 2013). The *F-tests* for the exogenous regressors are above 10 for both reduced-form equations, suggesting that the instrument is not weak.

Discussion

Our findings show that each stage of the 2PM is associated with distinct risk factors, but—at

the same time—local economic factors are a relevant dimension in influencing both stages. As for stage 1, namely the occurrence of online hate, economic insecurity—measured by the share of workers who perceive their job to be precarious—is a significant predictor. This links to the ongoing debate on the high level of job insecurity characterising the Italian labour market and its effects on people's discontent (Boeri and Brandolini, 2005; Modena, Rondinelli and Sabatini, 2014; OECD, 2018), supporting existing findings on its role in promoting discontent. Also, the transmission of social norms promoting inequality through family values acts as a risk factor for the occurrence of online hate. This result builds on the existing literature showing the importance of the family in shaping local social outcomes (Bertocchi et al., 2019; Duranton et al., 2009) including intolerance (Wilkinson et al., 2017), by adding evidence on its role in fostering online resentful behaviours. Finally, we find a small but significant association between cyberhate and voting for right-wing parties in agreement with existing works. This result seems to support the persistency of resentment, showing that it also manifests itself in the everyday conversations on social media alongside at the ballot box as detailed in the literature on the geography of EU discontent (Dijkstra, Poelman and Rodríguez-Pose, 2019; Iammarino, Rodriguez-Pose and Storper, 2019; Martin et al., 2018).

The results of the second stage of the 2PM, intensity of hate tweets, provide the evidence referring to our main research question, showing the relevant role of income inequality in fuelling cyberhate intensity. The estimates from the CF-TSRI model show that the influence of income inequality on the intensity of cyberhate holds when we account for the potential endogeneity bias for income inequality, which could arise due to sorting of households to accommodate their preferences towards the public bad of intolerance. Our estimated effect

for income inequality on cyberhate contributes to existing literature detailing how income inequality is a risk factor for violent discriminative behaviours and anti-immigrant attitudes (Burgoon et al., 2019; Elgar et al., 2013; Engler and Weisstanner, 2020; Kunst et al., 2017; Wilkinson and Pickett, 2017), adding evidence on its effect on cyberhate. By showing the effect of inequality on a share of resentment expressed on the internet, it also contributes to the literature about the spatial effects of inequality on discontent (Iammarino et al., 2019; McCann, 2020; Rodríguez-Pose, 2018). In this respect, our evidence shows that more unequal places suffer from more cyberhate, with no evidence on the effect of real and perceived inequality between places. The latter aspect deserves a proper investigation, that goes beyond the scope of the present paper, so to contribute to the debate on the 'tale of two inequalities' (Rodríguez-Pose, 2020) and measuring the relative strength of interpersonal inequality and territorial inequality on cyberhate.

While our results on the direct effect of inequality seems in line with the literature, we also find that higher income inequality relates to cyberhate through its interaction with the local share of educated people, as shown by the positive and significant coefficient of the interaction term. This surprising finding can be explained referring to extant evidence depicting relevant country heterogeneity in the observed effect of education on hate (Finseraas, Skorge and Strøm, 2018; Lancee and Sarrasin, 2015; Weber, 2020). It can also be referred to other evidence highlighting that in some countries inequality triggers increased intolerance among higher educated people prompting the fear of losing status (Decelles and Norton, 2016; Jetten et al., 2017; Kunstman, Plant and Deska, 2016; LeBlanc, Beaton and Walker, 2015). Our finding also parallels recent evidence from Italy on the strong negative stereotypes towards immigrants observed in the educated group of teachers (Alesina et al., 2018) and to the fact

that Italy has a sizeable level of income inequality compared to other countries.

Additionally, we provide two other interesting findings. First, we find that immigrants appear to be not relevant in shaping cyberhate. Second, crime does not emerge as a significant risk factor. These findings are consistent with hate narratives that do not relate hate to actual crime or actual foreign population, but rather define it as an expression of resentment whose motivations reside elsewhere (Glaeser, 2005). Our evidence on foreign population refers to findings from the literature on the geography of EU resentment, which details that the actual geography of migration does not appear to be a strong player in shaping voting for anti-EU parties (Dijkstra, Poelman and Rodríguez-Pose, 2019). By showing that the actual outlook of immigrants do not relate to hate tweets against minorities, our findings also provide quantitative support to the experimental game evidence showing that inequality triggers fears capable of altering people's perception of different ethnicities, to the point of believing that minority ethnic groups are far larger than they actual size (Krosch, Tyler and Amodio, 2017; Kunst et al., 2017). Finally, our empirical investigation seems to suggest that trust and civic engagement do not exert any meaningful influence on cyberhate (detailed results can be found in Supplementary Tables A.4 and A.10). These results do not allow, however, us to conclude that social capital is ineffective in countering cyberhate, given the many dimensions in which it can be measured. A detailed investigation of the cyberhate/social capital nexus goes beyond the scope of the present paper, but it appears to deserve further investigation.

Overall, our findings improve the understanding of the real-world local determinants of digital resentment, contributing to the literature on the effect of the economic dimension on intolerant behaviours (Gerstenfeld, 2017; Hall, 2013; Stephan et al., 2000). They also suggests that while the perception of job insecurity is enough for people to start tweeting hate, inequality is needed for hate tweets to increase in their volumes.

Postestimation diagnostics and robustness checks

The robustness of our results is assessed through several postestimation diagnostics and checks.

To account for a broad range of potential spatial triggers for intolerant behaviours we have considered different control variables, checking for their robustness as predictors by progressively including them into the 2PM model specification and subsequently testing the goodness-of-fit of the resulting estimates. Several tests have been performed to detect the most meaningful control covariates to be included in the final 2PM and CF-TSRI specifications (likelihood ratio tests, Wald tests, contrast test, Akaike's information criteria and Bayesian information criteria have been used to detect the potential core control variables; then, we have considered the postestimation diagnostics of all the regressions where the potential core control variables were fixed and combined with different sets of control variables).[7] We present the postestimation diagnostics for CF-TSRI model, since this model specification also allows to account for endogeneity of income inequality, hence conveying more information about the relationship between cyberhate and our considered independent variables. Notably, the probit model of the stage 1 is the same for the CF-TSRI and the 2PM specification, since it has no endogeneity issue relating to income inequality to be addressed. The results from the probit model hold also after removing the biggest LLMAs and with province fixed effects (Supplementary Table A.5), as well as when population is removed from the covariates due to its correlation with the share of hate tweets (Supplementary Table A.3, Column 3). Multicollinearity does not appear to be an issue (average variance inflation factor [VIF] = 1.44). The model performs well in terms

of goodness-of-fit and specification; it also displays outstanding discrimination in terms of sensitivity and specificity (see Supplementary Tables A.6 and A.7 and Supplementary Figure A.4). The log-linear model does not suffer from collinearity (average VIF = 6.91) or omitted variables (Ramsey test: *p*-value = 0.4640). We tested the interactions between the Gini index and family types alongside the interaction between the Gini index and rural area, but none was significant. Results hold when province fixed effects are included instead of regional fixed effects (Table 4, column 1).[8]

We include several robustness checks specifically targeting the zeros characterising our outcome variable. Our main specification assumes that the zero values are true zeros. However, the observed zero might be observations for which the potential outcome is latent (Dow and Norton, 2003). To account for this, we estimate an SSM (Cameron and Trivedi, 2010) and discriminate between the 2PM and the SSM by assessing which model has the strongest predictive power (Dow et al., 2003; Madden, 2008; Santos-Silva et al., 2015). Our exclusion restriction for the SSM is given by the share of population above 70 years old over digital natives. In fact, ISTAT (2018) shows that nearly 91% of the population over 75 years old did not have access to the internet in 2017. Furthermore, Twitter statistics about the age profile of users in Italy highlights that the 96% of Twitter users are below 70 years old (Global Web Index, 2015). This variable satisfies the condition for being an exclusion restriction (Cameron et al., 2010),[9] since the age profile considered in this case has an effect on tweet production, differently form the age cohorts which are relevant for the design on the instrumental variable for income inequality. Through two-step procedure it is possible to get consistent and robust estimates (Greene, 2003; Leung et al., 1996; Wooldridge, 2010) with error clustered at spatial level (for a detailed discussion and results, see Supplementary Tables A.12 and A.13).

Comparing the predictive power between the 2PM and the SSM, the former is confirmed as the preferred specification (see Supplementary Table A.16), although results between the two models are similar. Alongside the nature of zeros in the outcome variable, also their magnitude represents a feature worth considering in choosing the proper modelling strategy. In this regard, results from the 2PM and the SSM are compared with results from a single-stage OLS model where the zeros are assumed not to be a cause for concern (see Supplementary Table A.14 and Supplementary Table A.15 for detailed estimates). The single-stage OLS displays the lowest predictive power among competing model specifications (see Supplementary Table A.16), but still confirms our main findings.

Finally, we also consider a different normalisation for our dependent variable, dividing by all the resident population in each LLMA. Again, our main results are confirmed (see Supplementary Table A.15 and A.17).

Conclusions

Cyberhate is spreading quickly raising concerns among researchers, policy makers and the general population. Its diffusion is prompting more investigations into its determinants. This paper empirically identifies a causal relationship between income inequality and cyberhate in the case of Italy. Income inequality is capable of producing cyberhate through both a direct and an indirect effect. The latter linked the local share of educated people.

Our study has interesting implications for both theory and policy. From a theoretical perspective, our findings are consistent with, and contribute to, the literature acknowledging that cyberhate does not happen in a spatial vacuum (Hall, 2013), also adding cyberhate to other intolerant attitudes that are determined by economic inequality (Martin et al., 2018; Rodríguez-Pose, 2018; McCann, 2020; Wilkinson et al., 2017). Further research could

Table 4. *Estimates for the CF-TSRI approach applied to the stage 2 of the 2PM in several robustness checks (province fixed effects in column 1; removing outlier in column 2).*

CF-TSRI 2PM stage 2: Structural-form equation (y = share of online hate)			
	(1) Province FE coef		(2) No outlier coef
Educated people	0.379***	Educated people	0.238**
	(0.102)		(0.119)
Gini index	9.724***	Gini index	7.320***
	(2.836)		(1.701)
Educated people × Gini index	3.784***	Educated people × Gini index	2.638***
	(0.997)		(0.069)
Incomplete stem	6.902***	Incomplete stem	1.943***
Family (*Non-Egalitarian*)	(1.045)	Family (*Non-Egalitarian*)	(0.420)
Stem Family (*Non-Egalitarian*)	6.644***	Stem family (*Non-Egalitarian*)	1.444**
	(0.802)		(0.637)
Communitarian	—	communitarian	3.045**
Family (*Egalitarian*)		family (*Egalitarian*)	(1.333)
Reduced eq residuals	−1.055***	Reduced eq residuals	−3.648**
(direct effect)	(0.284)	(direct effect)	(1.695)
Reduced eq residuals	−1.395***	Reduced eq residuals	−1.596**
(interaction)	(0.112)	(interaction)	(0.698))
Controls	*YES*	Controls	*YES*
Observations	524	Observations	523
Province FE	*YES*	Regional FE	*YES*
R squared	0.588	*R* squared	0.454

Control: offline hate, marginalised area
CF-TSRI 2PM stage 2: Reduced-form equations

a: direct effect (y = Gini index)

Educated people	0.043***	educated people	0.0423***
	(0.005)		(0.004)
Synthetic Gini index	0.544**	Synthetic Gini index	1.029***
	(0.014)		(0.264)
Educated people × Gini index	0.096	Educated people × Gini index	0.182
	(0.156)		(0.116)
Controls	*YES*	Controls	*YES*
Observations	524	Observations	523
Province FE	*YES*	Regional FE	YES
R squared	0.559	*R* squared	0.462
F-stat ex instrument	12.04	*F*-stat ex instrument	15.14

b: interaction (y = educated people × Gini index)

Educated people	0.083**	Educated people	0.068*
	(0.027)		(0.038)
Synthetic Gini index	1.377	Synthetic Gini index	2.631
	(1.840)		(2.030)
Educated people × Gini index	0.416***	Educated people × Gini index	0.324***
	(0.051)		(0.063)
Controls	*YES*	Controls	*YES*
Observations	524	Observations	523
Regional FE	*YES*	Regional FE	YES
R squared	0.173	*R* squared	0.0746
F-stat ex instrument	11.02	*F*-stat ex instrument	10.52

Control: offline hate, marginalised area, family types.
Standard errors in parentheses are clustered at province level (column 1) and regional level (column 2). ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

explore how different aspects of income inequality influence cyberhate, for instance by considering the effects of income distances between the affluent, the middle class and the poor. Another interesting avenue for further research could be assessing another dimension of inequality, namely territorial inequality, to contribute to the research on the relative influence of interpersonal inequality and spatial inequality.

We also find that persistent social norms pivoting around inequality and self-centredness play a role in the creation of cyberhate. Local social norms are capable of influencing resentful behaviours alongside other relevant socioeconomic features already identified by the literature (Bertocchi et al., 2019; Duranton et al., 2009). Additionally, we also provide evidence on the role of economic insecurity acting as a risk factor for online hate, relating to recent findings identifying a relationship between economic insecurity and populism (Guiso et al., 2017). Overall, the findings of the paper provide further support to the association between economic anxiety and behaviours against minority groups.

Our findings have implications for the thriving policy debate on countering cyberhate. The evidence provided in the paper clearly shows that policy initiatives aiming at reducing the level of income inequality can contribute to countering social anxiety, as expressed through cyberhate. Hence, our findings provide support to the policy approaches which do not focus entirely on banning content publication which collides with freedom of expression rights. Moreover, addressing the factors that trigger cyberhate would allow to overcome the several shortcomings which are affecting the existing bulk of policy initiatives. Currently, the main policy focus is on improving commitment and accountability of internet service providers, social media platforms and digital users (ElSherief et al., 2018). This approach suffers from prosecutions vagaries, that dampen the deterrence effect, and from the cheap and fast relocation of hate contents in more favourable jurisdictions that prevents prosecution (McGonagle, 2013). Another policy implication suggested by our results is that, in Italy, people's perception of social injustice has not been countered by the existing social programs, aligning with existing analysis on the effectiveness of existing policy frameworks to counter resentment (Iammarino et al., 2019).

We note the limitations of this study. Clearly our findings refer to Italy, hence there are issues regarding their generalisation to other contexts. Future research should broaden the geographical scope of the analysis to assess whether inequality acts as a determinant for online hate more widely. Moreover, the focus of our investigation is online hate generated and diffused through Twitter. Other social networks such as Youtube and Facebook, as well as the dark web, represent relevant online media through which digital hatemongers operate, although being characterised by an architecture which does not easily allow for extensive geo-tagged data extraction.

Supplementary material

Supplementary data are available at *Cambridge Journal of Regions, Economy and Society* online.

Endnotes

[1] http://www.voxdiritti.it/ecco-le-mappe-di-vox-contro-lintolleranza/.

[2] https://users.humboldt.edu/mstephens/hate/hate_map.html.

[3] In the robustness checks, we estimate another normalisation given by hate tweets per capita.

[4] Some covariates that have not proved to be significant are omitted for readability issues: robberies, homicides, thefts, usury, and gambling machines per inhabitants.

[5] For a detailed description of the design of instrument, see Valderrama and Rodriguez-Castelan (2019) and Enamorado et al. (2016).

[6] We have tested other interactions for income inequality: trust, foreign population, crime perception of job insecurity to find non significance.

[7] More details are provided in Supplementary Tables A.4 and A.9–A.10, and Supplementary Table A.18 also shows the correlation matrix.

[8] The results are also robust to the exclusion of outliers. In fact, the data show a mild outlier corresponding to the LLM '*Petralia Sottana*', located in Sicily, where the geo-located dataset of online hate tweets allocates a surprisingly high number of tweets (see also Supplementary Figure A.5). The estimation of the log-linear fitted model removing the outlier do not change the results, as displayed in Table 4, column 2.

[9] Other potential exclusion restrictions have been tested, such as broadband connectivity and 3G connectivity, but they do not satisfy the necessary conditions.

Acknowledgements

References

Abreu, M. and Öner, Ö. (2020) Disentangling the Brexit vote: the role of economic, social and cultural contexts in explaining the UK's EU referendum vote, *Environment and Planning A: Economy and Space*, **52**: 1434–1456.

Acciari, P. and Mocetti, S. (2013) *Una mappa della disuguaglianza del reddito in Italia, Banca d'Italia*. Roma: Questioni di economia e finanza. 208.

Alesina, A. et al. (2018) *Revealing Stereotypes: Evidence from Immigrants in Schools*. Cambridge, MA: NBER Working Paper Series.

Alesina, A., Brioschi, B. and La Ferrara E. (2021) Violence against women: a cross-cultural analysis for Africa. *Economica*, **88**: 70–104.

Anderson, D. M., Crost, B. and Rees, D. I. (2020) Do economic downturns fuel racial animus?, *Journal of Economic Behavior and Organization*, **175**: 9–18.

Andersen, R. and Fetner, T. (2008) Economic inequality and intolerance: attitudes toward homosexuality in 35 democracies, *American Journal of Political Science*, **52**: 942–958.

Andriani, L. and Sabatini, F. (2015) Trust and pro-social behaviour in a process of state capacity building: the case of the Palestinian territories, *Journal of Institutional Economics*, **11**: 823–846.

Assimakopoulos, S., Baider, F. and Millar, S. (2017) *Hate Speech in the European Union: A Discourse-Analytic Perspective*. Berlin: SpringerBriefs.

Bansak, K., Hainmueller, J. and Hangartner, D. (2016) How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers, *Science*, **354**: 217–223.

Baum-Snow, N. and Ferreira, F. (2015) Causal inference in urban and regional economics. In G. Gilles Duranton, J. V. Henderson, and W. C. Strange (eds) *Handbook of regional and urban economics*, 3–68. Amsterdam: Elsevier.

Bernatzky, C., Costello, M. and Hawdon, J. (2021) Who produces online hate? An examination of the effects of self-control, social structure, & social learning, *American Journal of Criminal Justice*. doi:10.1007/s12103-020-09597-3

Bertocchi, G. and Bozzano, M. (2019) *Origins and implications of family structure across Italian Provinces in historical perspective, Studies in Economic History*. In C. Diebolt et al. (eds). *Cliometris and the Family: global patterns and their impact on diverging development*, pp. 121-147. Berlin: Springer Verlag.

Billing, C., McCann, P. and Ortega-Argilés, R. (2019) Interregional inequalities and UK sub-national governance responses to Brexit, *Regional Studies*, **53**: 741–760.

Boeri, T. and Brandolini, A. (2005) The age of discontent: Italian households at the beginning of the decade, *Giornale degli Economisti*, **63**: 449–487.

Boustan, L., Ferreira, F., Winkler, H. and Zolt, Eric M. (2013) The effect of rising income inequality on taxation and public expenditures: evidence from U.S. Municipalities and school districts, 1970–2000, *Review of Economics and Statistics*, **95**: 1291–1302.

Burgoon, B. et al. (2019) Positional deprivation and support for radical right and radical left parties, *Economic Policy*, **34**: 49–93.

Burnap, P. and Williams, M. (2014) Hate speech, machine classification and statistical modelling of information flows on Twitter: interpretation and communication for policy decision making. *Policy & Internet*, **7**:1–18. doi:10.1002/poi3.85

Cameron, A. C. and Trivedi, P. K. (2010) *Microeconometrics Using Stata: Revised Edition*. College Station, TX: StataPress.

Castells, M. (2001) *The Internet Galaxy: Reflections on the Internet, Business, and Society*. New York, NY: Oxford University Press, Inc.

Cavaille, C. and Marshall, J. (2019) Education and anti-immigration attitudes: evidence from

compulsory schooling reforms across Western Europe, *American Political Science Review*, **113**: 254–263.

Chan, J. (2019) The effect of college education on intolerance: evidence from Google search data, *Applied Economics Letters*, **26**: 83–86.

Cheng, Z. et al. (2011) Exploring millions of footprints in location sharing services. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

Citron, D. K. and Norton, H. (2011) Intermediaries and hate speech: Fostering digital citizenship for our information age, *Boston University Law Review*, **91**: 1435–2131.

Costello, M. and Hawdon, J. (2018) Who are the online extremists among us? Sociodemographic characteristics, social networking, and online experiences of those who produce online hate materials, *Violence and Gender*, **5**: 55–60.

Côté, S., House, J. and Willer, R. (2015) High economic inequality leads higher-income individuals to be less generous, *Proceedings of the National Academy of Sciences of the United States of America*, **112**: 15838–15843.

Decelles, K. A. and Norton, M. I. (2016) Physical and situational inequality on airplanes predicts air rage, *Proceedings of the National Academy of Sciences of the United States of America*, **113**: 5588–5591.

d'Hombres, B. and Nunziata, L. (2016) Wish you were here? Quasi-experimental evidence on the effect of education on self-reported attitude toward immigrants, *European Economic Review*, **90**: 201–224.

Dijkstra, L., Poelman, H. and Rodríguez-Pose, A. (2019) The geography of EU discontent, *Regional Studies* **54**: 1–17.

Dow, W. H. and Norton, E. C. (2003) Choosing between and interpreting the Heckit and two-part models for corner solutions, *Health Services & Outcomes Research Methodology*, **4**(5): 5–18.

Duggan, M. (2017) Online harassment 2017, *Pew Research*, 1–85.

Duranton, G., Rodríguez Pose, A. and Sandall, R. (2009) Family types and the persistence of regional disparities in Europe, *Economic Geography*, **85**: 23–47.

Dustmann, C. and Fasani, F. (2016) The effect of local area crime on mental health, *Economic Journal*, **126**: 978–1017.

Elgar, F. J. et al. (2013) School bullying, homicide and income inequality: a cross-national pooled time series analysis, *International Journal of Public Health*, **58**: 237–245.

ElSherief, M. et al. (2018) Hate lingo: A target-based linguistic analysis of hate speech in social media. *Proceedings of the International AAAI Conference on Web and Social Media*, **12**: 42–51.

Enamorado, T. et al. (2016) Income inequality and violent crime: evidence from Mexico's drug war, *Journal of Development Economics*, **120**: 128–143.

Engler, S. and Weisstanner, D. (2020) The threat of social decline: income inequality and radical right support, *Journal of European Public Policy*, **28**: 153–173.

Eurobarometer (2016) *Media Pluralism and Democracy – Special Eurobarometer 452, Digital Single Market*. doi:10.2838/007658

European Commission (2018) *Third Evaluation of the Code of Conduit on Countering Illegal Online Hate Speech*. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#monitoringrounds

Eurostat (2019) *S80/S20 Income Quintile Share Ratio by Sex and Selected Age Group*. https://ec.europa.eu/eurostat/databrowser/view/ilc_di11/default/table?lang=en

Facebook (2019) *Facebook Transparency*. Menlo Park: Facebook.

Finseraas, H., Skorge, Ø. S. and Strøm, M. (2018) Does education affect immigration attitudes? Evidence from an education reform, *Electoral Studies*, **55**: 131–135.

Gagliardone, I. et al. (2015) *Countering Online Hate Speech, UNESCO Series on Internet Freedom*.

Gerstenfeld, P. B. (2017) *Hate Crimes: Causes, Controls, and Controversies*. Thousand Oaks, CA: SAGE.

Gest, J., Reny, T. and Mayer, J. (2018) Roots of the Radical Right: Nostalgic Deprivation in the United States and Britain, *Comparative Political Studies*, **51**: 1694–1719.

Glaeser, E. L. (2005) The political economy of hatred, *Quarterly Journal of Economics*, **120**: 45–86.

Global Web Index (2015) *GWI Social Q1*. https://insight.gwi.com/hs-fs/hub/304927/file-2615393475

Görzig, A., Milosevic, T. and Staksrud, E. (2017) Cyberbullying victimization in context: the role of social inequalities in countries and regions, *Journal of Cross-Cultural Psychology*, **48**: 1198–1215.

Greene, W. (2003) *Econometric analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Green, D. P., McFalls, L. H. and Smith, J. K. (2001) Hate crime: an emergent research agenda, *Annual Review of Sociology*, **27**(1): 479–504.

Guiso, L. et al. (2017) *Populism: Demand and Supply*. Milan: Bocconi IGIER working paper No. 610.

Hainmueller, J. and Hopkins, D. J. (2014) Public attitudes toward immigration, *Annual Review of Political Science*, **17**: 225–249.

Hall, N. (2013) *Hate crimes*, 2nd ed. Abingdon: Routledge.

Hawdon, J. et al. (2020) The enthymemes of supporting President Trump: explaining the association between structural location, supporting the president, and agreeing with online extremism, *Social Science Computer Review*, 1–18. doi:10.1177/0894439320905767

Himelboim, I., Mccreery, S. and Smith, M. (2013) Birds of a feather tweet together: integrating network and content analyses to examine cross-ideology exposure on Twitter, *Journal of Computer-Mediated Communication*, **18**: 40–60.

Hine, G. E. et al. (2016) Kek, Cucks, and God Emperor Trump: A measurement study of 4chan's politically incorrect forum and its effects on the Web. Proceedings of the Eleventh International AAAI Conference on Web and Social Media, (Icwsm), pp. 92–101.

Iammarino, S., Rodriguez-Pose, A. and Storper, M. (2019) Regional inequality in Europe: evidence, theory and policy implications, *Journal of Economic Geography*, **19**(2): 273–298.

ISTAT (2016) *Rapporto BES*.

ISTAT (2018) *Aspects of Daily Life: Internet – Age Detail. In Multipurpose Survey on Households: Aspects of Daily Life—General Part*.

Jetten, J. et al. (2017) "Fear of falling": economic instability enhances collective angst among societies' wealthy class, *Journal of Social Issues*, **73**: 61–79.

Kaakinen, M. et al. (2018) Social capital and online hate production: a four country survey, *Crime, Law and Social Change*, **69**(1): 25–39.

Kennedy, J. (2019) Populist politics and vaccine hesitancy in Western Europe: an analysis of national-level data, *European Journal of Public Health*, **29**(3): 512–516.

Korndörfer, M., Egloff, B. and Schmukle, S. C. (2015) A large scale test of the effect of social class on prosocial behavior, *PLoS One*, **10**(7): 1–48.

Kowalski, R. M., Limber, S. P. and McCord, A. (2019) A developmental approach to cyberbullying: prevalence and protective factors, *Aggression and Violent Behavior*, **45**: 20–32.

Kraus, M. W., Côté, S. and Keltner, D. (2010) Social class, contextualism, and empathic accuracy, *Psychological Science*, **21**: 1716–1723.

Krosch, A. R. and Amodio, D. M. (2014) Economic scarcity alters the perception of race, *Proceedings of the National Academy of Sciences of the United States of America*, **111**: 9079–9084.

Krosch, A. R., Tyler, T. R. and Amodio, D. M. (2017) Race and recession: effects of economic scarcity on racial discrimination, *Journal of Personality and Social Psychology*, **113**: 892–909.

Kunst, J. R. et al. (2017) Preferences for group dominance track and mediate the effects of macro-level social inequality and violence across societies, *Proceedings of the National Academy of Sciences of the United States of America*, **114**: 5407–5412.

Kunstman, J. W., Plant, E. A. and Deska, J. C. (2016) White ≠ Poor: Whites distance, derogate, and deny low-status ingroup members, *Personality and Social Psychology Bulletin*, **42**: 230–243.

Lancee, B. and Sarrasin, O. (2015) Educated preferences or selection effects? A longitudinal analysis of the impact of educational attainment on attitudes towards immigrants, *European Sociological Review*, **31**: 490–501.

Layte, R. and Whelan, C. T. (2014) Who feels inferior? A test of the status anxiety hypothesis of social inequalities in health, *European Sociological Review*, **30**: 525–535.

LeBlanc, J., Beaton, A. M. and Walker, I. (2015) The downside of being up: a new look at group relative gratification and traditional prejudice, *Social Justice Research*, **28**: 143–167.

Leung, S. F. and Yu, S. (1996) On the choice between sample selection and two-part models, *Journal of Econometrics*, **72**: 197–229.

Lick, D. J., Alter, A. L. and Freeman, J. B. (2018) Superior pattern detectors efficiently learn, activate, apply, and update social stereotypes, *Journal of Experimental Psychology: General*, **147**: 209–227.

Lowry, P. B. et al. (2016) Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model, *Information Systems Research*, **27**: 962–986.

Lu, R. and Sheng, Y. (2020) From fear to hate: How the Covid-19 pandemic sparks racial animus in the United States, *CEPR COVID Economics, Vetted and Real-Time Papers*, **39**:72–108.

Lunaria (2019) *Cronache di ordinario razzismo*. Rome: Lunaria.

Madden, D. (2008) Sample selection versus two-part models revisited: the case of female smoking and drinking, *Journal of Health Economics*, **27**: 300–307.

Manstead, A. S. R. (2018) The psychology of social class: how socioeconomic status impacts thought, feelings, and behaviour, *British Journal of Social Psychology*, **57**: 267–291.

Martin, R. et al. (2018) Globalisation at a critical conjuncture?, *Cambridge Journal of Regions, Economy and Society,* **11**: 3–16.

McCann, P. (2020) Perceptions of regional inequality and the geography of discontent: insights from the UK, *Regional Studies. Routledge*, **54**: 256–267.

McGonagle, T. (2013) The Council of Europe against online hate speech: conundrums and challenges. In Council of Europe Conference Expert Paper. Strasbourg: Council of Europe.

Modena, F., Rondinelli, C. and Sabatini, F. (2014) Economic insecurity and fertility intentions: the case of Italy, *Review of Income and Wealth*, **60**: S233–S255.

Müller, K. and Schwarz, C. (2018) Making America hate again? Twitter and hate crime under Trump, *SSRN*. doi:10.2139/ssrn.3149103

Musto, C. et al. (2015) CrowdPulse: a framework for real-time semantic analysis of social streams, *Information Systems*, **54**: 127–146.

Musto, C. et al. (2016) 'Modeling community behavior through semantic analysis of social data: the Italian hate map experience. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. ACM: 307–308.

Näsi, M. et al. (2017) Do routine activities help predict young adults' online harassment: a multination study, *Criminology & Criminal Justice*, **17**: 418–432.

OECD (2018) *OECD Employment Outlook 2018*. Paris: OECD Publishing.

OSCE-ODIHR (2010) *Incitement to Hatred vs. Freedom of Expression: Challenges of Combating Hate Crimes Motivated by Hate on the Internet*. Warsaw: OSCE Publishing.

OSCE-ODIHR (2019) *Hate Crime Reporting*. Warsaw: OSCE Publishing.

Perry, B. and Olsson, P. (2009) Cyberhate: the globalization of hate, *Information & Communications Technology Law*, **18**: 185–199.

Piff, P. K. and Moskowitz, J. P. (2017) *The Class–Compassion Gap*. In E. M. Seppälä et al. (eds). *The Oxford Handbook of Compassion Science*. Oxford: Oxford University Press.

Pinotti, P. (2015) Immigration Enforcement and Crime, *American Economic Review*, **105**: 205–209.

Rodríguez-Pose, A. (2018) The revenge of the places that don't matter (and what to do about it), *Cambridge Journal of Regions, Economy and Society*, **11**: 189–209.

Rodríguez-Pose, A. (2020) The rise of populism and the revenge of the places that don't matter, *LSE Public Policy Review*, **1**: 1–9.

Rooduijn, M. (2018) What unites the voter bases of populist parties? Comparing the electorates

of 15 populist parties, *European Political Science Review*, **10**: 351–368.

Santos-Silva, J. M. C., Tenreyro, S. and Windmeijer, F. (2015) Testing competing models for non-negative data with many zeros, *Journal of Econometric Methods*, **4**: 29–46.

Schmuck, D. and Matthes, J. (2015) How anti-immigrant right-wing populist advertisements affect young voters: symbolic threats, economic threats and the moderating role of education, *Journal of Ethnic and Migration Studies*, **41**: 1577–1599.

Sharma, S. (2015) Caste-based crimes and economic status: evidence from India, *Journal of Comparative Economics*, **43**: 204–226.

Silva, L. et al. (2016) Analyzing the Targets of Hate in Online Social Media, (March). http://arxiv.org/abs/1603.07709.

Stephan, W. G. and Stephan, C. W. (2000) An integrated threat theory of prejudice. In *Reducing Prejudice and Discrimination*, pp. 23–45. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Stock, J. H. (2001) Instrumental variables in statistics and econometrics. In N. J. Smelser and P. B. Baltes (eds). *International Encyclopedia of the Social & Behavioral Sciences*, 7577–7582. Amsterdam: Elsevier Science Ltd.

Suler, J. (2004) The online disinhibition effect. *Cyber Psychology & Behavior*, **7**: 321–326.

Sunstein, C. (2017) *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.

SWG (2017) La diffusione dell'odio in rete. Trieste: SWG edizioni.

Terza, J. (2017) Two-stage residual inclusion estimation: a practitioners guide to Stata implementation, *The Stata Journal*, **17**: 916–938.

Todds, E. (1990) *L'invention de l'Europe*. Paris: Seuil.

Twitter (2019) Twitter 15th Transparency Report. San Francisco, CA: Twitter Inc.

Valderrama D. and Rodriguez-Castelan C. (2019) Income in fewer hands, production by fewer firms: The effect of inequality on maket concentration. Society for the Study of Economic Inequality Conference Paper 348.

Weber, H. (2020) The educational divide over feelings about ethnic minorities: does more education really lead to less prejudice?, *Journal of Ethnic and Migration Studies*, 1–20. doi:10.1080/1369183X.2020.1810540

WHO (2018) *Assessment Report of the Global Vaccine Action Plan*. Geneva: Strategic Advisory Group of Experts on Immunization.

Wilkinson, R. G. and Pickett, K. E. (2017) The enemy between us: the psychological and social costs of

inequality, *European Journal of Social Psychology*, **47**: 11–24.

Wodtke, G. T. (2016) Are smart people less racist? verbal ability, anti-black prejudice, and the principle-policy paradox, *Social Problems*, **63**: 21–45.

Wolfe, R. M. (2002) Anti-vaccinationists past and present, *BMJ*, **325**: 430–432.

Wooldridge, J. (2010) *Econometric Analysis of Cross-Section and Panel Data*, 2nd ed. Cambridge, MA: The MIT Press.

Wooldridge, J. (2015) Control function methods in applied econometrics, *Journal of Human Resources*, **50**: 420–445.

Youtube (2019) Youtube Transparency. Mountain View: Google.