Year: 2024

Essays in Political Economy

MARCUCCI Andrea

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT D'ÉCONOMIE

Essays in Political Economy

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de

Doctorat en économie

par

Andrea MARCUCCI

Directeur de thèse
Prof. Dominic Rohner

Jury

Prof. Valérie Chavez-Demoulin, présidente
Prof. Sébastien Houde, expert interne
Prof. Juan F. Vargas, expert externe

Lausanne
2024

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT D'ÉCONOMIE

Essays in Political Economy

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de

Doctorat en économie

par

Andrea MARCUCCI

Directeur de thèse
Prof. Dominic Rohner

Jury

Prof. Valérie Chavez-Demoulin, présidente
Prof. Sébastien Houde, expert interne
Prof. Juan F. Vargas, expert externe

Lausanne
2024

UNIL | Université de Lausanne
Faculté des hautes études commerciales
Internef - 261
CH-1015 Lausanne

IMPRIMATUR

La Faculté des hautes études commerciales de l'Université de Lausanne autorise l'impression de la thèse de doctorat rédigée par

Andrea MARCUCCI

intitulée

Essays in Political Economy

sans se prononcer sur les opinions exprimées dans cette thèse.

Lausanne, le 20.08.2024

Professeure Marianne Schmid Mast, Doyenne

Members of the thesis committee

Prof. Dominic Rohner

Full Professor, University of Lausanne

Thesis Supervisor

Prof. Sébastien Houde

Associate Professor, University of Lausanne

Internal member of the thesis committee

Prof. Juan F. Vargas

Full Professor, University of Turin

External member of the thesis committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics

I hereby certify that I have examined the doctoral thesis of

Andrea MARCUCCI

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 5/7/24

Prof. Dominic ROHNER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Economics

I hereby certify that I have examined the doctoral thesis of

Andrea MARCUCCI

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Sébastien Houde Date: 06.07.2024

Prof. Sébastien HOUDE
Internal expert

University of Lausanne
Faculty of Business and Economics

PhD in Economics

I hereby certify that I have examined the doctoral thesis of

Andrea MARCUCCI

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 8th of July, 2024

Prof. Juan F. VARGAS DUQUE
External expert

Essays in Political Economy

Andrea Marcucci
University of Lausanne

2024

Acknowledgments

The PhD is a long journey, and, as with any long journey, you encounter many people who assist you along the way. First and foremost, I am deeply grateful to my supervisor, Dominic Rohner, for his generosity and guidance over the past five years. Without his support and insights, this accomplishment would not have been possible.

I am very thankful to my other coauthors, from whom I learned so much. Thank you, Marius, for the time you dedicated to me from the early stages of my PhD, with weekly meetings to set the foundation of our project, and for pushing it forward until now. You taught me how fortunate we are in doing our job. Thank you, Mathias, for your insightful suggestions and for showing me that there is always a deeper question to be asked. Thank you, Alessandro, for the trust you placed in me and the support you provided; having the opportunity to work with you has been a highlight of my PhD. Last but not least, I would like to thank Devis Decet, co-author of the third chapter of this thesis. Without your sharp mind, this thesis would be missing its best chapter. Thank you for being such a great co-author and friend. Special thanks also go to the members of my thesis committee, Sébastien Houde and Juan Vargas, for their insightful comments that will help me to improve substantially the papers in the future.

This journey would not have been the same without having amazing friends to share it with. Thanks to all the people in my Gerzensee cohort, especially the Lausanne crew, Brendan, George, Jeremias, Katja, Marco, Nathan and Teo, as well as Giuseppe, the best problem-sets group mate. A big thank you to my first office mates, Noemi and Jeremy, as well as Elio who joined us later, you were always there to help and encourage me from day one. I am grateful to my colleagues in the same field of research, Andrea, Resuf, Alice, Tiziano, Gema and Julian. Thank you for the insightful discussions and the fun we had together. A special thanks goes to the Italian crew: Giacomo, Enrico, Kevin, Fabrizio, Luca (and the others), you made home even closer than it is. I am grateful to the Swiss Romande people: Laia, Adrien, and Pauline. Even if I failed to learn French, you were always there to talk and hang out. Finally, thanks a lot to Guillaume and Sofia, you are amazing colleagues to have.

Before I started my PhD, someone convinced me that becoming an economist was not such a bad idea. All these people have supported me all the way here. I would like to thank my friends from the master in Bologna, Caterina, Eugenia, Michele and Riccardo, and my colleagues at JRC, Andrew and Matteo.

As in any journey, you need a solid basis to face the challenges you will encounter along the way. I would like to thank my family, who always supported me, and taught me to be curious,

and my friends from Genoa, who helped me grow as a person, especially Clara, Pietro and Alice, despite the distance you were always there for me in the hard times of this journey. I am grateful to my climbing friends, who made me possible to fully appreciate the beauty of this country and to engage in very different challenges than the academic ones in the Alps: Giacomo, Alberto, Matteo, Ludovico, Andrea, Nicola, Matteo and Mattheus.
Finally, and most importantly, I would like to thank Anna, the love of my life. Thank you for your constant support, for keeping the distance between us as small as possible all these years, for always believing in us. Sharing this experience with you has made the difficult moments easier and the successes worth achieving. You are the rock I can always rely on. Thank you for the joy you bring into my life every day, this journey would not have been possible without you by my side.

Contents

Introduction

Conflicts between and within countries are one of the most serious plagues of humanity, hindering the development of nations and the well-being of individuals. According to the Peace Research Institute Oslo (PRIO), 2023 was the most violent year since the end of the Cold War, with 122,000 battle-related deaths and a record number of 59 conflicts worldwide. Understanding the determinants of violence and studying the institutional frameworks that could reduce the likelihood of war and foster cooperation between social groups and countries is fundamental to promoting economic growth.

Democracies have been shown to be more effective than other institutional settings in reducing violence (see Acemoglu and Robinson 2000 and related work). In the first chapter of the thesis, "Ballot or Bullet: The Impact of the UK's Representation of the People Act on Peace and Prosperity" with Dominic Rohner and Alessandro Saia, we study the impact of voting rights extension on social violence, during the "Age of Reform" in the nineteenth-century Victorian era. Importantly, we focus on the democratizing effects of the Second Reform Act (1867), which almost doubled the number of enfranchised citizens by extending the voting right to all men, regardless of property ownership, excluding only the very poor who paid no taxes. We construct a novel dataset at the borough level with information about social violence generated by applying natural language processing techniques to newspaper articles. We leverage information on the pre-reform distribution of property rental values in each borough to predict the increase in eligible voters. In particular, we exploit exogenous variation around the pre-reform property qualification to vote using it to instrument for the number of new eligible voters within each borough after the reform. We find a strong pacifying effect of democratization on social violence. The average increase in the number of eligible householders generated by the reform roughly halves the average risk of social violence in the sample.
Looking at the mechanisms at work, we show that the two main drivers of violence reduction are civic engagement and local economic growth. With respect to the former, we find that the reform led to more competitiveness in the political arena, reducing the probability of re-election for incumbents and increasing the proposal of petitions from civil society. For the latter, exploiting the presence of job advertisements in local newspapers, we show that the average increase in the number of voters resulting from the reform increased economic activity by about 15.5%. Overall, we show that democratization processes reduce violence and stimulate economic growth.

In the second chapter, "The Years of Lead: Terrorism and Voting Behavior in the Italian 1970s" I take a complementary approach by trying to understand the impact of political violence and terrorism on voting preferences and, as a consequence, on political institutions. During the 1970s Italy experienced a season of extreme political violence perpetrated by both left and right-wing groups, the so-called "Years of Lead". During this period, about 1600 people were killed or severely injured in acts of extreme political violence. By digitizing new electoral data at the polling station level, as well as granular data related to terrorist events, I examine the impact of violence on individuals' voting preferences in three major Italian cities (Milan, Rome, and Turin). Exploiting within-city variation in the intensity of violence and voting preferences, I am able to control for a variety of confounding factors that could make terrorist attacks endogenous to political preferences. Using a difference-in-differences identification strategy at the local level, comparing precincts located close to terrorist attacks with those located farther away, I show that an additional episode of political violence widens the vote gap between the incumbent party (Christian Democracy) and the main opposition party (Communist Party), almost doubling it. This rally-around-the-flag effect persists for at least three electoral terms and is even more substantial when more severe event types are considered. The effect seems to be driven mainly by public security concerns. Using data from the Manifesto Project, which assigns scores to the policies proposed by parties' political manifestos, I show that an additional terrorist attack is associated with a substantial increase in votes for all parties supporting stricter law and order measures. These results suggest that terrorism, by creating a climate of fear and insecurity in the population, can increase support for incumbent parties by generating a rally-around-the-flag effect.

In the third chapter of my dissertation, "Water Wars", with Devis Decet, I turn to investigate the relationship between climate change and conflict. Climate change is one of the most significant global challenges currently faced by humanity. Between 2023 and 2050, it is expected to cause approximately 250,000 additional deaths per year. Over the past two decades, numerous studies (see Burke et al., 2015 for a review) have demonstrated the link between climate shocks and violence, highlighting this relationship as an additional factor contributing to the deadly effects of climate change. In our paper, we investigate the impact of droughts on conflicts over water resources in the African continent. Water is fundamental for human life and economic activity, especially in a rural context like the African one, where there exists a lack of infrastructure for water management. Moreover, the presence of a multitude of different ethnic groups can increase cooperation costs in sharing resources when they become scarcer. By constructing a novel dataset containing fine-grained information about water presence and the rivers' network structure, we show that droughts happening downstream increase

conflicts in upstream water-rich locations. This evidence suggests that climate shocks push individuals to seek access to water upstream along the rivers' network, where they can exert more control over the resource. This specific mechanism helps to shed new light on previous findings in the literature related to the spillover effects of climate shocks on conflict. Our main estimates suggest that a drought downstream increases the likelihood of severe conflict events happening in upstream water-rich locations by about 7% with respect to the average conflict incidence.

Confirming our hypothesis that higher cooperation costs are an important factor leading to fight for resources, we show that our main coefficient of interest is stronger in those areas where water is more unevenly distributed among different ethnic groups. Moreover, we find that the economic value of water matters. Indeed, we identify a stronger effect of our main coefficient in those areas of the continent with presence of agriculture.

Climate change does not impact populations only through more periodic droughts, but also via longer-term desertification processes. We show that in those locations where water presence has been decreasing in the past 40 years, populations are more inclined to fight over the resource as a consequence of climate shocks. In terms of policy implications, in the last part of the chapter, we stress how countries with better institutions, such as more democratic states or lower levels of corruption, seem to be better equipped to avoid conflicts over water resources.

A common side effect of conflict is displacement of people who flee their countries to seek safety and asylum protection. In the last chapter of my thesis, "Price and Prejudice: Housing Rents Reveal Racial Animus" with Marius Brülhart, Gian-Paolo Klinke, Dominic Rohner and Mathias Thoenig, we investigate the reaction of hosting communities toward asylum seekers. According to the International Organization for Migration (UN), the number of international migrants has grown from 84 million globally in 1970 to 281 million in 2020, with a total of 26.4 million refugees. Migrants' arrival is not always accepted by the local populations, leading to the rise in consensus for extreme right-wing parties, protests, and social grievances. As an example, in a recent Eurobarometer survey (2019), more than one-third of the respondents ranked immigration as the most important issue facing the EU (34%), way ahead of the second item, climate change (24%).

In our paper, we measure racial animus of the local population towards asylum seekers in the Swiss context. Using a difference-in-differences strategy, we examine the impact of the opening and closing of cantonal asylum centers on the rents of nearby housing units. Importantly, we have information about the nationality of origin of the refugees living in a given center. Given the quasi-random allocation of asylum seekers to the centers, this allows us to disentangle the impact of the nationality composition of the refugees living in a given center. We find

that the opening of a center is on average associated with a drop of 3.8% in rental prices of housing units in proximity to the center. When looking at differences in terms of asylum seekers nationality of origin we find that the drop in local rental prices in the vicinity of centers populated by above-median shares of asylum seekers from Sub-Saharan Africa equals 4.9%, whereas around centers with below-median shares of Sub-Saharan Africans it equals 2.7%. This result holds when controlling for a proxy of asylum seekers' crime propensity based on the nationality of origin. We interpret our findings as evidence of taste-based discrimination from the local communities. Looking at the characteristics of the resident population, we find suggestive evidence that neighborhoods with a more educated local population or greater genetic similarity to the refugees housed in the centers exhibit less discrimination.
The results of the study suggest that exists a "price of prejudice" which is a burden not only for the objects of discrimination but also for the subjects. This implies that policies promoting tolerance can generate a sort of double dividend, on top of reducing grievances in the society.

In summary, this thesis sheds new light on the complex interplay between institutions, violence and economic development. The first two chapters use historical data to explore the two sides of the relationship between institutions and conflict. Chapter 1 demonstrates that democratization processes reduce social violence and promote economic growth. In contrast, Chapter 2 reveals that terrorism can increase support for incumbent parties by creating a climate of fear in the population. On the other hand, the last two chapters address current global challenges related to the causes and consequences of conflict. Regarding the causes, Chapter 3 shows that climate shocks exacerbate conflicts over vital water resources in Africa. Finally, Chapter 4 focuses on one of the many consequences of war: the migration of asylum seekers. In particular, we examine the reactions of host communities to the arrival of refugees, highlighting the economic consequences of racial animus and the social costs of discrimination. Taken together, these findings have substantial policy implications, highlighting the importance of robust institutions, inclusive policies, and climate change mitigation in fostering peaceful societies and economic development.

CHAPTER 1

Ballot or Bullet: The Impact of the UK's Representation of the People Act on Peace and Prosperity

Andrea Marcucci University of Lausanne
Dominic Rohner University of Lausanne
Alessandro Saia University of Bologna

Abstract

Does democracy hold its promise of curbing domestic political violence? While the matter has been heatedly debated for decades, so far not much reliable causal evidence exists. To study this question, we focus on the United Kingdom's Victorian Age of Reform, and in particular the Representation of the People Act of 1867 – which is widely regarded as a critical juncture in the history of democratization. Constructing a novel borough ("urban center") level dataset on social conflict events and economic performance around the 1868 Elections (the first elections where newly enfranchised citizens could vote), we exploit arguably exogenous variation in enfranchisement intensity. We find a strong and robust peace-promoting effect of franchise extension and identify as a major channel of transmission the increase of the population's political influence (voice) and local economic growth.

Keywords: Social Violence, Social Conflict, Riots, Democracy, Enfranchisement, Franchise Extension, Voting, Elections, Voice, Growth, Prosperity, Development.

JEL: C33, D72, D74, N43, O17.

"This is why I say it's the ballot or the bullet. It's liberty or it's death. It's freedom for everybody or freedom for nobody." Malcolm X
"The Reform Act of 1867 was one of the decisive events, perhaps the decisive event, in modern English history. It was this act that transformed England into a democracy (...)." Gertrude Himmelfarb

1. Introduction

The extension of voting rights has often been used by elites as a way to stave off civil unrest. This phenomenon is well-documented in political economics research (see Acemoglu and Robinson, 2000 and related work).[1] However, as later explained, empirical evidence in this domain is lacunary at best. Most of the existing literature explaining the relationship between voting rights and conflict lacks exogenous variation and is confined to correlational and cross-country evidence, and risks being biased by various confounders. Recent advances in Optical Character Recognition (OCR) and Natural Language Processing (NLP) combined with the availability of comprehensive digital newspaper archives now make it possible to carry out fine-grained within-country investigations, allowing our research to address the caveats of past findings and uncover causality.
In an attempt to address this literature gap, the current paper focuses on England's particularly decisive "Age of Reform" in the 19th century Victorian era. Suffrage extensions from this period have not only shaped modern Britain but have also inspired the rise of liberal democracies around the world. Among the series of notable reforms, the 1867 UK's Representation of the People Act, also referred to as Second Reform Act, has resulted in the largest relative surge in enfranchised voters – roughly doubling the electorate at the time. In addition to being of great historical importance, this reform act implemented a countrywide property value threshold to determine voter eligibility. Our research exploits this arguably exogenous threshold in the identification strategy. In a nutshell, the turbulence of this key historical period, together with the exogenous variation available and the newly assembled data, offers an opportunity for novel evidence on the aforementioned classic research question.
To study the impact of this reform act on social conflict, we have built a novel panel dataset of borough level monthly data which includes new measures of social conflict, local economic growth, and several other control variables. Prior to 1867, voting rights were related to residential property values - a clear proxy for income and wealth. In particular, men owning

[1] One strand of the literature sees democracy as commitment device, with democratization leading to redistribution in favor of the population, thereby reducing reasons for revolt (see Acemoglu and Robinson, 2001, Acemoglu and Robinson, 2005, Fearon, 2011, and Bidner, Francois and Trebbi, 2014). In contrast, in another strand of the literature, democracy reduces asymmetric information and through this channel curbs the risk of conflict (Laurent-Lucchetti, Rohner and Thoenig, 2019). Both types of settings yield the prediction that democratization reduces the likelihood of conflict.

a property with a rental value of at least ten pounds were eligible for voting. The reform extended suffrage to all men, irrespective of the rental value of their property or whether they owned property but excluded the very poor, defined as those not paying taxes (i.e. the so called "poor rates"). We leverage information on the pre-reform distribution of property rental values in each borough to predict the increase in eligible voters. More specifically, we draw on the arguably exogenous idiosyncratic variation around the previous voting threshold of a housing rent value of ten pounds. Instrumenting the extent of franchise extension using this idiosyncratic variation, we estimate how enfranchisement mattered for the risk of social conflict. We detect a strong and significant pacifying effect from suffrage extension. Quantitatively, the average increase produced by the reform (i.e. roughly doubling the number of electors) reduced likelihood of social conflict by over 12 percentage points, which amounts to about half of the baseline conflict risk. Expressed in terms of standard deviations, an enfranchisement change of one standard deviation (roughly 40% more voters) resulted in a decreased risk of social conflict of about 0.13 standard deviations. Interestingly, we find that it is not at the moment of the *de jure* passing of the act that it deploys its effects but only when the *de facto* change in representation has materialized. This is quite intuitive, as registration offices where struggling to register all new voters (Davis and Tanner, 1996), hence making it natural for newly eligible voters to have residual doubts on the process before the elections actually took place.

When investigating the specific mechanisms at work, we find that enfranchisement has led to a substantial rise in the population's political influence (which we will refer to as *voice*) and state accountability. In particular, we find that the participation of new voters in the 1868 elections has made UK politics more competitive and has increased elected representative turnover. This strengthened civic engagement was also reflected by a substantial increase in the number of petitions received by the House of Commons. Importantly, drawing on novel data from job advertisements, we also detect a strong and significant impact of franchise extension on boosting local economic growth, in particular in areas with large market potential. In terms of magnitude, increasing the number of electors by 89% (the average increase produced by the reform) is associated with an increase in economic activity of around 15.5%. Our findings indicate that more inclusive, pluralistic political institutions may foster economic growth, coinciding with recent findings of Acemoglu, Naidu, Restrepo and Robinson (2019) and Abeberesey, Barnwalz, Chaureyx and Mukherjee (2020), which in turn provides a fertile breeding ground for peace (in line with e.g. Miguel, Satyanath and Sergenti, 2004; Dell, Jones and Olken, 2014; König, Rohner, Thoenig and Zilibotti, 2017).

One caveat of the analysis is that it is inherently short-run, which makes it difficult to study mechanisms and channels of transmissions that take substantial time to materialize. In particular, our results are unable to detect any increase in state capacity and public spending linked

to increased enfranchisement: this is unsurprising given that such changes are only likely to materialize in the long run.

The current paper is related to several different areas of research. First, an empirical literature establishes correlations between democracy and social conflict at the country level. These studies fail to plausibly demonstrate causal links as they do not exploit exogenous variation in democratization. In addition, these papers find overall contradictory results. In particular, Fearon and Laitin (2003) find no significant effect of democracy, while Besley and Persson (2011*a*) conclude that the conflict-fuelling effect of negative shocks is muted by cohesive institutions. There is also evidence that the relationship between democracy and conflict is non-linear: Hegre, Ellingsen, Gates and Gleditsch (2001) conclude that full democracies and full dictatorships are associated with a lower conflict risk than intermediate regimes, Collier and Rohner (2008) find that democracy is linked to a lower conflict risk in rich countries while there is no beneficial effect in poor countries and Cervellati and Sunde (2013) detect a peace-promoting impact of "third wave" democratization mostly when transitions are non-violent. Furthermore, Fetzer and Kyburz (2018) conclude that local elections in Nigeria lead to more cohesive institutions that limit distributional conflicts between groups. In terms of potential pitfalls, Esteban, Morelli and Rohner (2015) stress that while consolidated democracy may be associated with fewer mass killings, initial democratization can induce an increased risk of violence. This finding is consistent with evidence from case studies documenting how ill-managed democratic transition instigates spikes of nationalist conflict (Snyder (2000), Mann (2005)). Likewise, Fergusson, Querubin, Ruiz and Vargas (2020) find that narrow elections of previously excluded left-wing parties in Colombia leads to backlash from right-wing paramilitaries. Also related to this type of literature, Collier and Vicente (2014), Cederman, Gleditsch and Hug (2013) and Dercon and Gutiérrez-Romero (2012) study episodes of (post-)electoral violence.

Another relevant branch of literature studies the drivers of franchise extension, including namely Acemoglu and Robinson (2000), Lizzeri and Persico (2004), Llavador and Oxoby (2005), Przeworski (2009), Doepke and Tertilt (2009), Aidt and Jensen (2014), Aidt and Franck (2015).[2] These works closely relate to more theoretical writings analysing how and why democracy reduces the scope for conflict. The democracy-peace nexus can be explained by the following mechanisms: the ability of the population to select tax rates and potentially enhance redistribution (see e.g. Acemoglu and Robinson, 2001 and the literature on democracy as commitment device), the capacity of democracy to reduce asymmetric information

[2]Our study of the effects of a massive, universal franchise extension for all ethnic groups (almost doubling in our case the number of eligible voters) is also complementary to the research on the effects of more specific laws or measures improving the representation of given minority groups (see Fujiwara (2015), Mueller and Rohner (2018), Facchini, Knight and Testa (2020), Lacroix (2020)).

(Laurent-Lucchetti et al., 2019), or strengthened state accountability under democracy (Collier and Rohner, 2008) – which may in turn improve state governance and boost the economy, thereby reducing grievances and increasing the opportunity cost of rebellion. Yet, democracy may also provoke adverse effects, by exacerbating electoral competition (see discussion above on the dangerous transition to democracy and on (post-)electoral violence), as well as galvanizing subversive uprisings through freedom of assembly.

In light of these potential channels and mechanisms, it is also important to mention the series of papers studying other implications of franchise extension and democracy, i.e on public finances (Aidt, Dutta and Loukoianova, 2006, Aidt, Daunton and Dutta, 2010), on economic growth (Acemoglu et al., 2019, Abeberesey et al., 2020), or on health outcomes (Besley and Kudamatsu, 2006, Kudamatsu, 2012). Furthermore, the current contribution is also part of the e conomic history literature studying Victorian England (Aidt et al., 2010, Berlinski, Dewan et al., 2011, Berlinski, Dewan and Van Coppenolle, 2014, Aidt and Franck, 2015, Aidt and Franck, 2019, Chapman et al., 2020).

Lastly, this paper is part of the emerging literature analysing how institutions and policies are able to reduce the scope of conflict, using arguably exogenous policy variations (see the recent literature survey of Rohner, 2022). Other contributions in this line of research have focused for example on the impact of food aid (Nunn and Qian, 2014), education (Rohner and Saia, 2019), grand coalitions (Mueller and Rohner, 2018) or reconciliation ceremonies (Cilliers, Dube and Siddiqi, 2016).

In a nutshell, the novel contribution of this paper is twofold: First, moving beyond correlational evidence at the country-level, it studies the impact of a very large-scale and arguably exogenous franchise extension on social conflict, drawing on newly constructed fine-grained social conflict data. Second, this more granular, newly assembled data allows us to investigate the mechanisms at work that link enfranchisement to civic peace.

The remainder of this paper is structured as follows. Section 2 describes the historical context of the UK's Second Reform Act. Section 3 describes the data used, Section 4 lays out the identification strategy and Section 5 presents the baseline results and a series of robustness tests. Section 6 is dedicated to a discussion of the main mechanisms at work. Finally, section 7 concludes. An extensive (Online) Appendix contains further detailed explanations and additional results.

2. Historical Context

Britain's 19th century has often been referred to as the "Age of Reform", during which several franchise extensions have made politics increasingly inclusive.[3] A first milestone was the 1832 "First Reform Act", which extended suffrage to the middle class. This reform act was introduced to appease (violent) popular demands for voting rights (with the 1831 "Queen Square" riots in Bristol being a famous example of such popular unrest). However, as the popular elite wanted to maintain political power, the electoral law only allowed men owning (housing) property worth at least £10 to vote, which precluded the working class from enfranchisement.

Thus, political protest continued, mostly fuelled by the urban and unfranchised working class (Harrison, 1962). After additional attempts to restore peace with insufficient reforms, the next major advance in enfranchisement was the "Second Reform Act" (known formally as the Representation of the People Act of 1867) granting voting rights to the urban male labour class in England and Wales for the first time. While introduced by the Conservative government under Prime Minister Benjamin Disraeli, it benefitted from support beyond party lines. It received Royal Assent by the British Crown on August 15, 1867. The bill was to be implemented in progressive stages over the following years as decided by the UK Parliament. The population's newly found political participation culminated in the 1868 elections, where incumbent Conservatives led by Prime Minister Benjamin Disraeli were defeated by the Liberals, paving the way for the Liberal William Gladstone to become the new Prime Minister.

The 1867 reform bill abolished the £10 per year threshold and granted voting rights to all house owners or occupiers (householders) in English and Welsh boroughs that had resided during the last 12 months and paid their taxes (i.e. the so-called "poor rates"). While there is no clear-cut rent threshold for householders being subject to paying the poor rates tax, the typical threshold may lie somewhere around £4.[4]

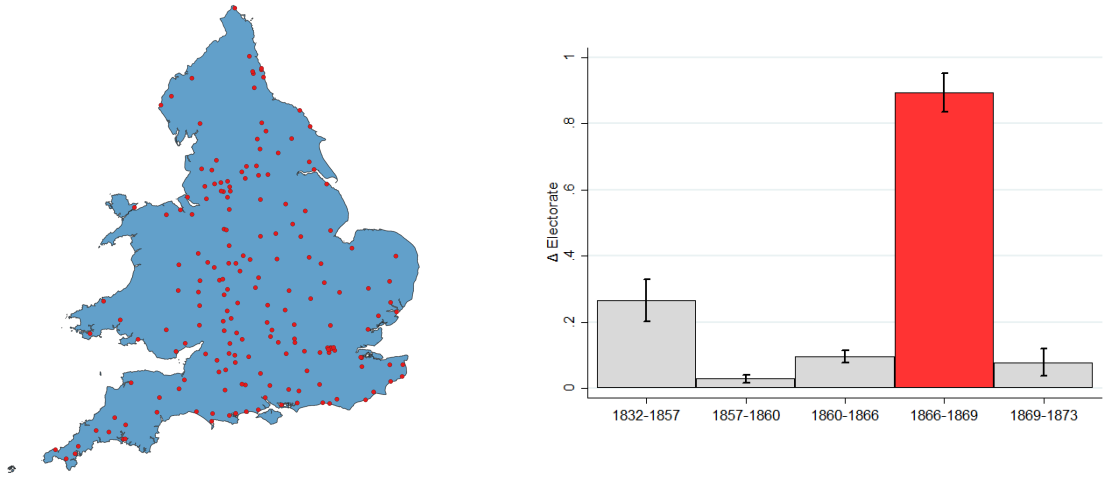
This 1867 reform took place in a climate of intense debate between Progressives, who wanted to extend suffrage to the greater population, and Conservatives, who were afraid of any radical extensions, that would confer too much political power to the working class that they perceived as untrustworthy (Zimmerman, 2003). In the end, the Representation of the People Act of 1867 was one of the most decisive reforms in English history (Himmelfarb,

[3] The following description draws on the accounts of Himmelfarb (1966), Smellie (1968), Fraser (1976), Davis and Tanner (1996), Zimmerman (2003), Schlager (2004), Lizzeri and Persico (2004), Saunders (2007), Lawrence (2009), Aidt et al. (2010), Berlinski et al. (2011), Turner and Zhan (2012), Berlinski et al. (2014), Aidt and Franck (2015), Chapman et al. (2018), Aidt and Franck (2019), Chapman et al. (2020), www.parliament.uk, the "Encyclopedia Britannica" and the "St. James Encyclopedia of Labor History Worldwide".

[4] Additional details on the Reform and on the rationale for using £4 as threshold for taxation (and hence voting rights) in the empirical analysis are provided in the Appendix A.

1966) and resulted in a path-breaking electoral reform: men who occupied or owned a house in urban areas and who paid taxes (the so called “poor rates”) were enfranchised, which roughly doubled the electorate in England and Wales from one to two million men. Panel B in Figure 1 emphasizes that the 1867 reform was a particularly important milestone in the history of UK electoral reform as it massively increased the numbers of newly enfranchised voters.[5] This paper exploits the 1867 "Second Reform Act” to study the effect of franchise extension on the likelihood of political unrest.[6]

Figure 1: British Boroughs and Electorate growth over 19th Century



(A) Boroughs included in the sample

(B) Electorate growth

NOTE: Panel A: Map of England and Wales with each dot representing a given borough included in the sample. Panel B: Each column displays the growth rate in the number of electors in British boroughs over different elections. Black lines indicate the 95% confidence interval. Electoral data are taken from official accounts of the parliamentary papers (House of Commons (1857), House of Commons (1866*a*), House of Commons (1869*b*) and House of Commons (1874*b*)). The timing of measurement of the electorate is not evenly spaced (hence, over longer time spans population growth could affect electorate growth) and does not always correspond to election years.

3. Data

We build a novel panel dataset at the borough [urban area] and month level. The baseline monthly dataset covers 184 boroughs of England and Wales over the period 1868-1869 (24

[5] One relevant question to ask is why there was not more backlash of members of the elite against this reform. One can think of several reasons. First of all, the Second Reform Act still contained some minimum wealth requirements for being able to vote. Hence, the poorest parts of the population were still excluded from politics, attenuating the elite’s fear of radical redistribution. Further, opposing openly a new act that had been voted in Parliament and had the support of both major parties (liberals and conservatives), as well as of the Prime Minister and the Queen would have been a risky endeavor.

[6] In the years after our sample period the UK government also tightened rules on firearms (e.g. with the 1870 Gun License Act and the 1903 Pistols Act).

months around the 1868 elections, the first elections after the passage of the Second Reform Act).[7] In our analysis we focus on boroughs that did not experience changes in their enfranchisement status, i.e. those that were enfranchised both before and after the reform and for which we exploit the "intensive margin" of numbers of individuals able to vote.[8] For robustness checks we also construct an extended longer time-span dataset, as well as a more fine-grained dataset at the weekly frequency.

3.1 Social Conflict Data

The dependent variable of interest we study is social conflict outcomes. In particular, in the baseline specifications we focus on the extensive margin of social conflict events, relying on a dummy variable taking a value of 1 in a given borough and month where at least one social conflict event took place and 0 otherwise. In the robustness checks, we also investigate the intensive margin and construct a variable of social conflict intensity at the borough and month level.

To construct our novel social conflict data, we start from the British newspaper archive and perform data scraping of a sample of over 300,000 newspaper articles containing at least one social conflict-related keyword.[9] This sample of newspaper articles stems from pieces published over our sample period by over 500 national or local newspapers. Our algorithm codes as social conflict observation news reports containing social conflict-related keywords linked to a given borough location. For illustration, the five newspapers with widest national cover of social conflict events for our sample period where London Evening Standard, The Sun, The Scotsman, Morning Post and London Daily News. The full list of newspaper sources is reported in Appendix F.[10]

One potential issue with social conflict data generated using historical newspaper articles is that factors such as media pressure (i.e. originating from other timely topics needing coverage) could affect reporting (see e.g. Jetter, 2017; Durante and Zhuravskaya, 2018). To

[7] In Panel A of Figure 1 are mapped all boroughs included in the sample.

[8] Importantly, applying this criterion implies that 87 percent of the boroughs are included in the sample (184 out of 211). The few missing ones are boroughs that either before or after the reform were disenfranchised (e.g. as sanction for mismanagement). The main reason that led to our choice is the fact that we are unable to retrieve the number of electors for the disenfranchised boroughs after the passage of the act and the number of electors for the enfranchised boroughs before the passage of the act. This is due to the fact that in those cases information on the numbers of electors is provided at a different level of aggregation over time. For example, for a disenfranchised borough, after the passage of the act the number of electors is only available at the level of the county constituency.

[9] The list of keywords used in the baseline analysis are Disturbance, Disturbances, Unrest, Riot, Riots, Rioters, Rioting, Tumult, Tumults, Disorder, Disorders. For more details see Appendix D.21.

[10] To provide –for the purpose of illustration– a few examples of social conflict-related events covered by our data, consider e.g. riots in Ashton-under-Lyne, Stalyridge, Bristol, Cardiff, Bolton, North Shields, among many others. See also some examples of newspaper articles in Appendix Figure A2.

address this, in the Appendix B.3 we show that social conflict data generated from our historical newspapers correlates strongly with data from an established existing social conflict dataset provided from an external source (Holland, 2005), and recently used e.g. by Caprettini and Voth (2020) (see Appendix Table A3). Potential reporting bias could also arise if media owners have vested interests, as far as enfranchisement is concerned. Thankfully this concern is attenuated by the very short sample period over which media ownership is unlikely to change.

3.2 Electoral Data

The main explanatory variable is the number of enfranchised citizens. The first election taking place after the passing of the Second Reform Act in 1867 is the 1868 United Kingdom general election [17 November – 7 December 1868]. The number of voters in a given borough is time-varying over our sample period, corresponding to the pre-reform number of voters until October 1868, and the post-reform voter numbers from November 1868 onwards. The number of enfranchised electors before and after the reform are taken from official accounts of the parliamentary papers (House of Commons, 1866*a* and House of Commons, 1869*b*).[11] Note that in a robustness analysis we move the analysis from the monthly to the weekly level (Appendix D.14), which allows to take into account that the elections only started in mid-November. We have also assembled additional voting statistics and electoral data from House of Commons (1866*b*), House of Commons (1869*a*), House of Commons (1874*a*) and Craig (1977). Finally, we have also collected data for other elections before and after 1868 (drawing on elections 1865 and elections 1874 data from House of Commons (1857), House of Commons (1866*a*) and House of Commons (1874*b*)), which we use for a placebo analysis.[12]

3.3 Instrumental Variable

As discussed in more detail below, we instrument for the number of newly enfranchised voters after the reform by exploiting idiosyncratic variation at the borough level in the structure of rents paid by householders. As later argued in further detail, when controlling for the average rents and inequality of rents, any remaining variation around the £10 threshold can be seen as quasi-exogenous. Put differently, if two boroughs have the same average rents and rent inequality, but one has for some reason a higher share of rents below the pre-reform £10

[11] In particular, we use information on the "total number of voters on register of each borough" available in the various Parliamentary Papers. This value corresponds to "the number of electors on the register in force at the time of the election" available in the book "British Parliamentary Election Results 1832-1885" (Craig (1977)).

[12] Our main focus is on the 1868 United Kingdom general election [17 November – 7 December 1868], that are the first election after the passing of the Second Reform Act (August 1867). In the placebo analysis we use the 1865 general election [11–24 July 1865] and the 1874 general election [31 January – 17 February 1874].

cut-off (close to the threshold), this borough will (exogenously) experience a larger increase in new voters. To construct this instrument and the corresponding control variables, we draw on the fine-grained rents distribution data from House of Commons (1866*a*) (see Appendix B.2 for further information on the rents data).

3.4 Other Data

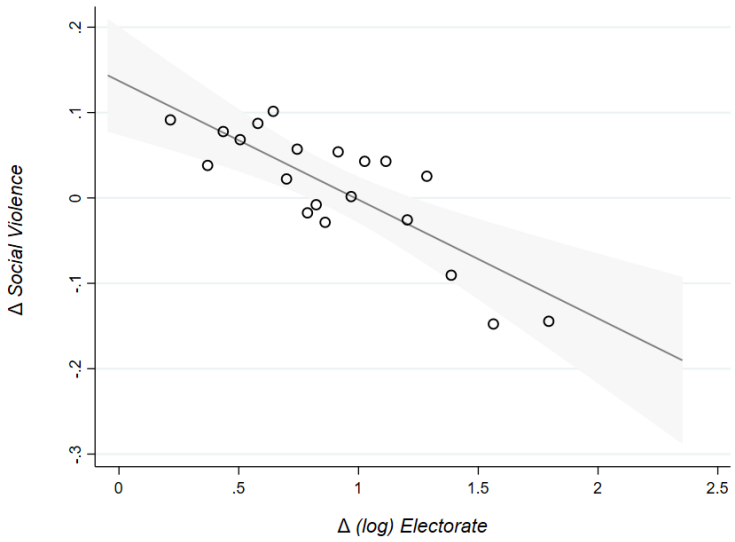
A series of control variables are included in the baseline and robustness specifications, namely average gross estimated rents and Gini index of rents, both computed using data from House of Commons (1866*a*), as well as population variables from House of Commons (1866*a*). Additionally, other borough-level employment-based variables (i.e. the share of working population, the share of elementary occupations, and the gender-ratio) have been constructed using the 1861 Population Census conducted by the Secretary of State of the United Kingdom (IPUMS, 2020). We have also constructed novel time-varying data on job ads at the borough level (as described in detail in Appendix B.4).

3.5 Descriptive Summary Statistics and Raw Association

The summary descriptive statistics are displayed in Appendix B. In particular, Table A2 depicts the key moments of the main variables of the analysis. Among the 4,416 borough-months in our sample roughly 23 percent experienced social conflict. As far as the increase in the electoral base is concerned, during our Second Reform Act period of interest (1866 to 1869) the number of eligible voters almost doubled, while over the other periods concerned (before and after) average franchise extension was in the order of magnitude of between 3 to 27 percent. The summary statistics of various further variables reveal additional interesting patterns, e.g. that public spending did not increase between 1868 and 1869, which is in line with the discussion of mechanisms below. Moreover, Appendix B.2 provides further information on rents, such as the distribution of rents in Figure A1. Table A64 lists average rents, rent inequality (measured using the Gini Index), as well as the number of householders in total and per relevant rent bracket for each borough.

Figure 2 below provides a bin scatter illustration of the raw association between enfranchisement and social conflict reduction. While this negative correlation between the changes in electors and social conflict is visually striking, it could be partially driven by confounders. Hence, in the remainder of the paper, we study in much more depth how franchise extension affected social conflict, drawing on arguably exogenous variation in the level of enfranchisement.

Figure 2: Raw association between enfranchisement and change in social conflict



NOTE: Bin scatter graph with each bin summarizing values of several boroughs. The red line represents the linear prediction plot based on the underlying data. The x-axis depicts the Δ in the (log) electorate pre and after the Second Reform Act observed in a given borough, while the y-axis plots the difference in (average) social conflict between the period before and after November 1868.

4. Model Specification and Identification

4.1 OLS Specification

In the goal of identifying the impact of franchise extension on social conflict, we perform a difference-in-differences analysis exploiting the variation of the total number of electors over time and across boroughs. We start with the following specification for the OLS regressions:

$$Social\ Violence_{it} = \beta_0 + \beta_1\ (log)\ Electorate_{it} + FE_i + FE_t + \epsilon_{it}, \tag{1}$$

where the variable $Social\ Violence_{it}$ is a dummy that takes a value of 1 if a violent event was observed in a borough i and during month t. $Electorate_{it}$ represents the number of registered electors at the time of the election observed in borough i during period t.[13]

Our identifying variation originates from the fact that the intensity of new electors enfranchised in the Elections of 1868 varies widely across boroughs. The specification features borough fixed effects (which filter out time-invariant borough characteristics such as e.g. el-

[13]Note that given that we control for borough fixed effects, the estimates would be identical if we had our explanatory variable based on the share of the population registered as electors instead of the absolute number of registered electors.

evation, sea access, longitude and latitude) and monthly time dummies (which control for country wide shocks, such as e.g. major political and economic nationwide shocks). The standard errors are clustered at the level of the 184 boroughs in all regressions (unless indicated otherwise).

4.2 Instrumentation

The above specification has the advantage of filtering out time-invariant borough characteristics and nationwide shocks, but one may worry that borough-specific shocks and trends could confound with the coefficient of interest of franchise extension. In particular, it could be that poorer boroughs experience a bigger franchise extension and at the same time are subject to shocks or trends that foster "gentrification" and tackle social unrest. Any pacifying effect attributed to enfranchisement could hence be spuriously driven by borough-specific socio-economic changes and policies.

Another potential identification threat is related to the fact that our variable of interest corresponds to registered voters. Given the burden of registration procedures and the role of potential partisan and non-partisan interests in the registration process (see Davis and Tanner, 1996), it could be that the increase in enfranchisement may be affected by borough-level social dynamics. For example, one could imagine that in boroughs where poorer citizens have more political momentum, there is more registration (and hence enfranchisement) and more redistributive policies, reducing the scope for social unrest. Further, in more conservative boroughs with anti-enfranchisement views registration could be made deliberately complicated, hence limiting the number of new voters, while more liberal areas may embrace franchise extension wholeheartedly. Hence, a correlation in OLS between franchise extension and peace could be spurious and driven by the underlying momentum of liberal forces in a given borough.

To address such concerns, we run two-stage least square (2SLS) regressions where we instrument for the scope of enfranchisement by exploiting idiosyncratic –arguably as good as random– variation in the number of newly enfranchised householders for each borough. As discussed above, the reform led to the removal of the previous administrative threshold of £10 rental value. If there are two boroughs, A and B, with exactly the same average rents and same rent inequality, but for some idiosyncratic reason in borough A there is a slightly higher mass of citizens with rental value right below the previous £10 threshold, borough A will, for quasi-random reasons, experience greater enfranchisement than borough B. We control for average rents and rents variance to filter out potentially confounding effects of prosperity and inequality. Hence, in our instrumental variable strategy, the remaining identifying variation in mass around the £10 threshold can arguably be seen as good as random.

In particular, our instrumental variable (IV) is labelled *Eligible Householders*$_{it}$ and corresponds

to the eligible population in each period. In particular, in the period before the Election of 1868, the value of our variable *Eligible Householders* is equal to number of householders above £10, while in the period after the election it is equal to the number of householders living in a house with rental value above £4 (i.e., those living in a house with rental value between £4 and £10 who were previously banned from voting but enfranchised by the Second Reform Act plus the number of householders with rental value above £10 who could already vote before the reform).[14]

Figure 3 illustrates graphically the identifying variation of the first stage of our 2SLS estimation. We see that our instrument is strongly correlated to the increase in the number of electors in a given borough. Note that the raw correlation (Panel A) also holds for the residual correlation when controlling for average rents and inequality in rents (Panel B).[15]

As far as the exclusion restriction of the IV is concerned, the instrument is valid under the assumption that the share of householders newly eligible to be enfranchised by the Reform only affects the social conflict risk through the increase in the electorate and not through some other channel. The plausibility of this assumption is supported by the fact that we control for both the average level of rents and for the inequality of rents. Hence, given that these controls both account for the general level of wealth, prices and inequality of a given borough, the remaining variation in rents amounts to quasi-random idiosyncratic variation around the previous voting threshold of £10.

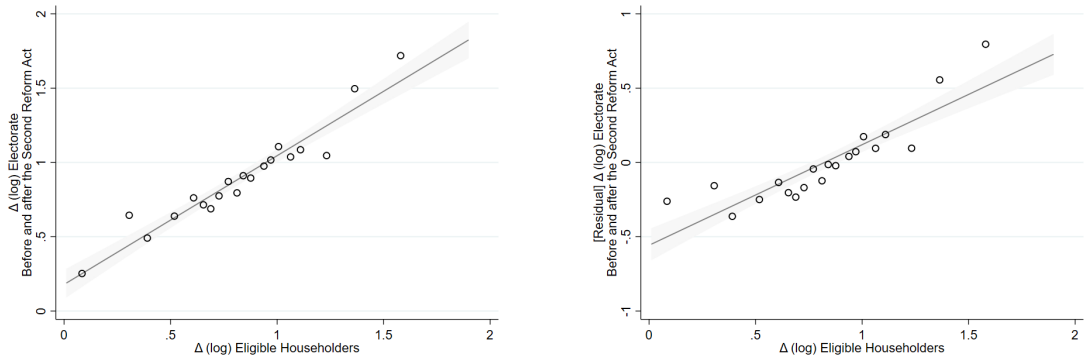
5. Main Results

Table 1 displays the main results. In column 1, we start with the OLS specification where we directly regress social conflict incidence on the number of electors. We include borough fixed effects and time dummies, which makes this specification a classic difference-in-differences setting. We find a sizeable effect of enfranchisement reducing social conflict which is statistically significant at the 1 percent level. Quantitatively, we find that increasing the number of electors by the average increase produced by the reform (89%), reduces the social conflict likelihood by over 12 percentage points, which amounts to about half of the baseline conflict risk. Expressed in terms of standard deviations, a one standard deviation change in enfranchisement (roughly 40% more voters) results in a decreased risk of social conflict risk of about 0.13 standard deviations. In column 2, we add average rents interacted with monthly time dummies, and, in column 3, we further include the interaction of rent inequality with monthly time dummies. It turns out that the coefficient magnitude remains very stable and the statisti-

[14] Note that in terms of terminology we use the expressions "householders above £10" and "rent value > 10" interchangeably (and analogously for other thresholds).

[15] In Appendix C we present the bin scatter illustration of the raw association between our instrument and social conflict reduction.

Figure 3: Δ (log) Electorate pre and after the Second Reform Act and Δ (log) householders below threshold



(A) Unconditional (B) With Controls

NOTE: Panel A displays the values of Δ (log) Electorate pre and after the Second Reform Act and Δ (log) householders below threshold in 184 English boroughs along two axes. Panel B displays the values of the residual of Δ (log) Electorate pre and after the Second Reform Act when control for average rents and inequality in rents and Δ (log) householders below threshold in 184 English boroughs along two axes.

cal significance high when controlling for the average and distribution of rents. Hence, while conceptually important to control for average prosperity and inequality, it turns out that for our sample the results are virtually unchanged when these controls are added. In columns 4-6 we estimate the reduced-form impact of our instrument on social conflict incidence. We find that the increase in eligible voter householders significantly reduces the potential for social conflict. This result holds in a specification controlling for the same batteries of fixed effects, averages, and inequality of rents as in the first three columns. Finally, in columns 7-9 we perform a two-stage least square (2SLS) estimation where our IV of the number of eligible householders is in the first stage used to instrument for the number of electors and in the second stage the estimated electors number is used as regressor on the dependent variable of social conflict incidence. We find that our instrument is a strong predictor of the number of electors (as pointed out by the F-stats that are substantially above the conventional threshold of 10) and that an increase in electors statistically significantly drives down the risk of social conflict.[16] It is equally important to note that the coefficient size of the 2SLS estimation is of a similar order of magnitude as in the OLS estimation, and is very stable across all 2SLS results (i.e. across columns 7-9). This may be consistent with the view that most *de jure* eligible new voters actually ended up complying (registering), and that registration biases and other endogenous confounders were rather limited.

[16]The strength of our instrument is also confirmed by the Olea-Pflueger robust test for weak instruments (Olea and Pflueger, 2013). The Effective F-statistic for column 7 is equal to 223 (and the critical value for the null hypothesis that the 2SLS bias exceeds 5 percent of the OLS bias is equal to 37.418).

Table 1: Democracy and Social Violence: Main Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.139***	-0.139***	-0.141***				-0.169***	-0.176***	-0.175***
	(0.0421)	(0.0446)	(0.0457)				(0.0562)	(0.0627)	(0.0620)
(log) Eligible Householders$_{it}$				-0.146***	-0.158***	-0.158***			
				(0.0483)	(0.0550)	(0.0549)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.389	0.392	0.397	0.389	0.392	0.396	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.237	.237	.237	.237	.237	.237	.237	.237	.237

Notes:The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.1 Robustness Analysis

In this subsection we briefly list the main robustness tests performed. In the interest of space, they have been relegated to the Appendix.

Additional Controls The first set of robustness checks focus on adding further control variables. In particular, for the purpose of addressing concerns on potential confounders, in Appendix D.1 a battery of additional socio-demographic control variables from the 1861 census are included (interacted with time fixed effects), while in the Appendix D.2, we include flexible functional form controls for rents brackets interacted with monthly time dummies, respectively. Appendix D.2 also shows that our rent bracket of interest between £4 and £10 is the only rent bracket that is a statistically significant predictor of conflict. Next, Appendix D.3 controls for fixed effects at the spatial NUTS3 times month level and Appendix D.4 accounts for electoral changes over the previous elections. Finally, Appendix D.5 includes alternative inequality measures. Our results prove robust to all these additional controls.

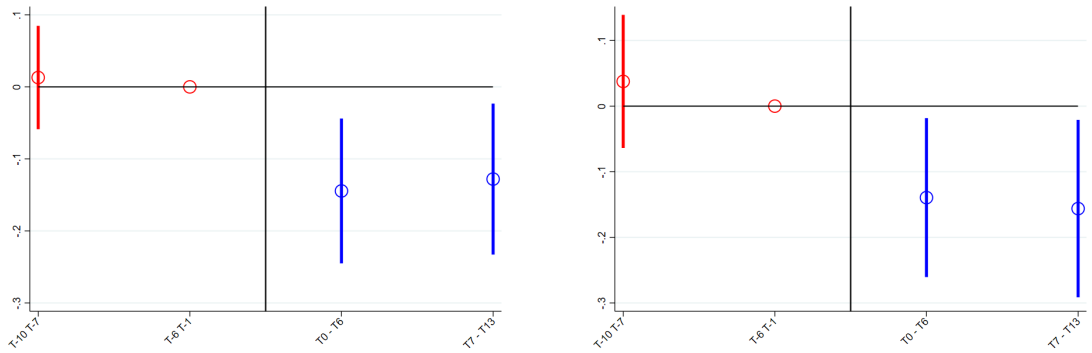
Alternative Instruments In a second set of robustness checks we investigate whether our findings are sensitive to the exact way of constructing our instrument. While in Appendix D.6 an alternative property threshold of £0 is considered (i.e. we include in the instrument householders for any rent band), Appendix D.7 reproduces our baseline results but drawing on earlier rent data (addressing concerns on rent manipulation). Our results remain very similar for all these specifications.

Alternative Dependent Variable To assess the scope of our findings, in Appendix D.8 we display the results for an alternative dependent variable, focusing not only on the extensive margin, but also on the intensity of social conflict incidence. We find that enfranchisement does not only affect the likelihood of social conflict, but also its intensity (e.g. measured by the number of social conflict events).

Estimation methods and inference The next set of robustness checks investigate whether our findings hinge on the exact statistical methodology applied, or hold across a broader range of methodological approaches. Our main dependent variable being a binary 0-1 dummy, we replicate our results using logit in Appendix D.9. Next, we focus on the issue of a common pre-trend. In Figure 4 below, we perform an event study. In particular, we split our sample period into 4 sub-periods in order to assess whether enfranchisement observed in the elections correlates with differences in social conflict behavior prior to the reform. We do not detect any pre-trend and the effect of enfranchisement only kicks in after the 1868 election. Similarly, Appendix D.10 contains an event-study analysis for boroughs with below- versus above-median enfranchisement. Going further, in Appendix D.11 we rely on the Synthetic Control Method (SCM), recently applied e.g. by Abadie and Gardeazabal (2003), Billmeier and Nannicini (2013), and Saia (2017), to guarantee –by construction– an identical pre-trend

before the enfranchisement reform. These sets of methodological sensitivity checks leave the conclusion from the baseline analysis unchanged: Franchise extension has curbed social conflict. Finally, in Appendix D.12 we perform alternative (two-way, spatial and more granular) clustering of standard errors, which allows to account for complex correlation patterns of standard errors. The statistical inference remains very similar.

Figure 4: Democracy and Social Violence: Leads and Lags



(A) Electorate (B) Eligible Householders

NOTE: The figure displays the coefficients of estimates of leads and lags of the variables *(log) Electorate Post*$_{it}$ (i.e. the (log) number of electors post-reform) in *Panel A* and *(log) Eligible Householders Post*$_{it}$ (i.e. the (log) number of eligible householders post-reform) in *Panel B*. In each of the two specifications we also control for *(log) Electorate Pre*$_{it}$ [*(log) Eligible Householders Pre*$_{it}$] (i.e. the number of electors [eligible householders] pre-reform) in *Panel A* [*Panel B*]. These variables are interacted with time dummies taking value 1 for each time-window. Time-windows are displayed on the horizontal axis (omitted period is [T-6:T-1]). Estimates are obtained including interactions of average rents and rent inequality with monthly time dummies. In this way we obtain leads and lags specifications comparable with columns 1 and 4 of Table 1. The dependent variable is $Social\ Violence_{it}$.

Outliers and different time units In this next set of sensitivity checks we focus on the sample composition. In particular, we assess robustness to the presence of potential outliers and to different time units. More specifically, in Appendix D.13 investigates whether the findings are driven by outliers. For this purpose, the regressions are re-run when boroughs, newspapers and random days are dropped from the sample, revealing that the results are very stable across specifications. Further, Appendix D.14 depicts the findings when the temporal unit of observation is either the week or the pre-post reform period. In all cases, our results prove robust to these sensitivity tests.

Time trends and extended sample period As mentioned above, a key identifying assumption is the existence of a common pre-trend. A complementary approach to the exercises already discussed above (and displayed in Figure 4 and in Appendices D.10 and D.11), is to control for a borough-level time trend and to extend the sample in order to allow for longer pre- and post-treatment periods. The corresponding results are presented in Table A25 in Appendix D.15. We find that our baseline results continue to hold when we include a borough-specific

linear trend, and when we extend the panel length up to two years before and after our baseline sample.[17] The extended sample allows us also to investigate the role of *de jure* enfranchisement (at the passing of the act) versus *de facto* political representation (at the first election where the newly enfranchised electoral body was called to vote). Similarly to the exercise presented in Figure 4 above, we divide the sample into several sub-periods to identify any differential effect after the elections versus after the passage of the act. The corresponding results are presented in the Appendix D.15. It is shown that the *de jure* passing of the law did not suffice to curb social conflict, and we also do not observe any change in social conflict events between the passing of the act and the first election thereafter in 1868. In contrast, it is found that after the 1868 election the level of social conflict decreases. This finding is not entirely surprising: given that voters had to register and that registration offices were overwhelmed and struggling to register all new voters (Davis and Tanner, 1996), it made sense for people at the moment of the passing of the act to have residual doubts on the process and on their ability to participate in the next vote. However, in 1868 when the elections actually took place, the promises of the reform were realized and confidence in the process and representation were consolidated.

Placebo analysis The next set of robustness tests address concerns about our identification strategy "mechanically" picking up something else, e.g. due to measurement error or reporting bias. We start off in Appendix D.16 with an assessment of whether our instrumental variable also affected social conflict in other time periods, where the reform did not apply and where accordingly we would not expect any effect. We find that for both pre-post 1865 and pre-post 1874 elections, reassuringly no effects of 1868 reforms were detected. In a similar vein, in Appendix D.17 the main analysis is replicated by using sport events instead of social conflict events as the dependent variable. As expected, the reform only affects the latter, and not the former. This attenuates concerns about reporting bias from newspaper reports affecting our results, as any mechanical bias should also affect reports on sport and not just politics. Moreover, in Appendix D.18 we randomly assign treatment in 1,000 placebo datasets with the same average social conflict likelihood as the "true" data, finding reassuringly that it would be extremely unlikely that our results were found "by chance". Last but not least, in Appendix D.19 we randomly permutate the rents paid by householders to investigate the validity of our instrument. Reassuringly, no effects are found for this fake data.

Data construction Finally, we also carry out a series of robustness checks with respect to the

[17] Note that the extended sample comes at a price, as for longer sample durations there is a higher risk of unobserved heterogeneity and confounding factors, such as institutional changes that have occurred (e.g, the Ballot Act and several public health bills introduced during the 1860s and 1870s). These concerns make it crucial to include time trends when analysing the extended sample. Still, despite these, the longer sample offers overall a less compelling identification strategy, which is why we focus in the baseline estimations on the shorter, more homogenous sample (24 months).

algorithm used to detect and geo-code social conflict events (see Appendix D.20) and the keywords used (see Appendix D.21). Importantly, as major robustness check, in Appendix D.22 we also re-construct the social conflict measure relying –instead of the bag-of-words approach (which could be sensitive to the exact terms included)– on machine learning techniques, using a lasso model. In particular, we hand-code 1,000 strings (sentences or sentence parts) as indicating the presence of social conflict or not. Of these, 900 strings are then used to train the machine learning algorithm that is next applied out-of-sample on the 100 remaining hand-coded strings (yielding an out-of-sample accuracy of 93%). We then apply this algorithm to the full set of strings of all newspaper articles to construct an alternative lasso-based dependent variable. Strikingly, the sample mean of the resulting social conflict measure is extremely close to that of our baseline variable, and when replicating our baseline regressions using this very different alternative data construction approach, we find very similar results (see Appendix Table A37). Further, in Appendix D.23 we apply a Natural Language Processing (NLP) approach to construct our social conflict variable. This sophisticated method not only takes into account keywords but also makes use of sentence structures. Our baseline results prove robust to all these various sensitivity tests on data construction. Finally, another major worry could be that our variable of social violence may not only capture important events of social conflict, but may also occasionally pick up smaller, very local events (e.g. bar rows). In order to address this concern, in Appendix D.24 we restrict the social violence variable construction to include only events that are so large-scale that they are also reported in newspapers with headquarters in boroughs far away from where a given event took place. We find that our results are –if anything– stronger when focusing on such bigger events.

6. Channels

After having scrutinized our results for a broad range of robustness checks in the previous section, we now study the underlying mechanisms and channels at work. In particular, we focus on the following three potential mechanisms through which enfranchisement potentially drives down the risk of social conflict: i) increased participation in the political arena [voice], ii) increase in economic activity, and iii) increased state capacity. We assess them successively. The main results are reported below, and supplementary robustness results on the channels and mechanisms are relegated to the Appendix E. Of course, a limit to our analysis is that these outcome variables and social conflict can be codetermined and there can be complex bi-directional causality links between e.g. social conflict and economic activity.

6.1 Voice: Increased Participation in the Political Arena and Beyond

A mechanism through which enfranchisement could deploy effects on social conflict outcomes is a higher state accountability, as greater political participation and inclusion may result in larger scrutiny. The key role of state accountability for reducing fighting has been stressed in Collier and Rohner (2008). To start with, one may expect –almost mechanically– that greater *de jure* eligibility for voting may result in a higher number of *de facto* voters casting their votes, which may affect the competitiveness of elections and political turnover. Beyond these effects, also general, non-electoral forms of civic engagement, such as petitions, may be affected. Interestingly, voting and petitions could be either complements or substitutes, as investigated below.

Number of Voters A determining pre-condition for this channel is that the newly enfranchised voters actually made use of their new-found powers and went voting, thereby increasing political openness and competition. In Table 2 we investigate this, running the following specification:

$$(log)\ Voters_{it} = \beta_0 + \beta_1 (log)\ Electors_{it} + FE_i + FE_t + \epsilon_{it} \tag{2}$$

We find that boroughs with more new eligible electors did indeed see a larger increase in the number of voters participating to the November 1868 elections.[18][19]

Contested Elections For more voters to trigger greater political accountability, a key question is whether this influx of new political actors has actually made the political arena more competitive. It turns out that this was indeed the case. As documented in Figure 5, in the elections of 1868 there has been a surge in political competition which is reflected as an increase in the number of contested elections.[20] Our findings are in line with the results of Berlinski et al. (2011) that highlight a rise in electoral competition following the reform act.

Political Turnover After investigating the notion of contested elections, we now question whether challengers have a real chance of defeating incumbents in the polls and if a healthy level of political turnover takes place in reality. This is an important pre-condition for effective citizen representation and would typically be violated if an established political elite holds a

[18] In Appendix E.1 we provide further evidence of the non-linear effect of enfranchisement on voters participation. Results provided in Table A43 suggest that boroughs with higher intensity of enfranchisement experienced a larger increase in the number of voters.

[19] One related question is whether the new voters did not only participate to elections, but on top of that were more active than previously-enfranchised citizens, thereby pushing upward the whole overall turnout. As reported in Appendix Table A44, the results on this are not very conclusive. The coefficient of interest is always positive but never statistically significant.

[20] Complementary to these findings, we also report in the Appendix (see Appendix Table A45) that new voters forming a plurality decreases significantly social violence and that the more newly enfranchised voters there are (and hence the more "voice" they have), the more social conflicts decrease.

Table 2: Voice - Democracy and Participation in the Political Arena

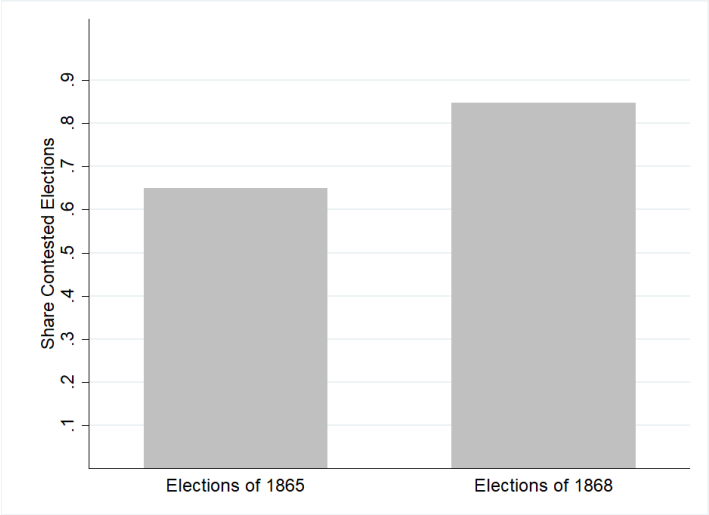
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: (log) Voters$_{it}$	OLS	OLS	RF	RF	2SLS	2SLS
(log) Electorate$_{it}$	1.102***	1.043***			1.226***	1.180***
	(0.105)	(0.0722)			(0.194)	(0.178)
(log) Eligible Householders$_{it}$			0.851***	0.826***		
			(0.152)	(0.150)		
Observations	216	216	216	216	216	216
R-squared	0.977	0.978	0.966	0.968	-	-
1st stage F-Stat	-	-	-	-	87	57
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*Rent Inequality$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes

Notes: The unit of observation is borough i and election t (where t corresponds to the elections of 1865 and 1868). OLS (2SLS) estimates are reported in columns 1-4 [5-6]. The dependent variable is the *(log) Voters$_{it}$* correspond to the electorate in a borough i observed in election t. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

de facto monopoly over political power and access to political office. A priori, the Second Reform Act offered a well-needed opportunity to "stir things up", allowing the entry of new faces into the political landscape. Below we investigate to what extent the enfranchisement has indeed favored political turnover and paved the way for new candidates to win elections. In effect, we digitalise new data from the British Parliamentary Election Results book, containing information about the name of all the candidates and elected politicians for the years 1832-1885. With the help of this data, we construct two new dependent variables: (i) a dummy equal to 1 if at least one incumbent candidate gets re-elected and (ii) the share of seats for which the incumbent party got re-elected. We estimate a difference-in-differences specification, regressing these dependent variables on our main variable of interest (log electorate), as well as borough and year fixed effects.[21] Our findings are displayed in Table 3. In

[21] For dependent variable (i) we also control for the number of seats in the town (which changed for some boroughs between the two elections), while for dependent variable (ii) there is no need to add this control, as we already express it as share of the number of seats.

Figure 5: Voice - Share Contested Election 1865-1868



NOTE: The two bars indicate the share of borough with contested elections in the elections of 1865 and 1868, respectively. Election data is obtained from House of Commons (1866*a*) and House of Commons (1869*b*).

the odd columns we display the results for the dependent variable (i), while the even columns refer to the dependent variable (ii). We find that suffrage expansion gave rise to an effective change in the political arena, reducing the likelihood that incumbent candidates and parties got re-elected.

Petitions One key element for accountable democratic societies is the relationship between the people and their politicians, and a vibrant civic culture typically gives birth to regular interaction that is not only limited to periodic elections. In particular, public petitions matter and are a powerful indicator of the extent to which "the people" have a "voice" in a given political system. There may be a complementarity between inclusive elections and further means of expression such as petitions – one could expect that enfranchisement and better representation of citizens may also lead to more vibrant civic exchanges between electors and the authorities. The right of the people of the United Kingdom to present petitions to the House of Commons dates to 1669, and the Nineteenth century is considered a "golden age" of parliamentary petitioning. In 1843, prime minister Benjamin Disraeli said, “I believe that at this moment the right of petition [. . .] is a more important and efficient right than has ever been enjoyed by the people of England in this respect” (Parliamentary Debates, 3rd Series, Vol 101, c673, 30 August 1843).

We use newly digitized archival records of the Select Committee on Public Petitions to retrieve all public petitions received by the House of Commons from 1868 to 1869. For each petition, we know where the petitions originated, the number of signatures related to the petition

Table 3: Voice - Democracy and Effects for Incumbents

	(1) OLS	(2) OLS	(3) RF	(4) RF	(5) 2SLS	(6) 2SLS
Dep. Variable:	*Incumbent Re-elected*$_{it}$	*Share Seats Inc. Party*$_{it}$	*Incumbent Re-elected*$_{it}$	*Share Seats Inc. Party*$_{it}$	*Incumbent Re-elected*$_{it}$	*Share Seats Inc. Party*$_{it}$
(log) Electorate$_{it}$	-0.351**	-0.105*			-0.327*	-0.190*
	(0.137)	(0.0623)			(0.186)	(0.106)
(log) Eligible Householders$_{it}$			-0.282*	-0.171*		
			(0.162)	(0.0934)		
Observations	366	366	366	366	366	366
R-squared	0.573	0.595	0.563	0.598	-	-
1st stage F-Stat	-	-	-	-	176	177
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The unit of observation is borough i and election year t. The sample covers 183 boroughs for the two elections in years 1865 and 1868. LPM (2SLS) estimates are reported in columns 1-4 [5-6]. The dependent variable in odd columns [1-3-5] is a dummy that takes value 1 if at least one incumbent gets re-elected in a given election t. The dependent variable in even columns [2-4-6] is the proportion of incumbent parties that get re-elected in a borough (i.e. $\frac{\#parties \quad re-elected}{\#seats}$) in a given election t. These dependent variables were constructed using data collected by the authors from Craig (1977). The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). All columns include Borough and Year of election, Average Rents by year and Gini Rents by year FEs. In odd columns [1-3-5] we also control for the number of seats in the town for each election (in even columns [2-4-6] our dependent variable is already divided by the number of seats). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and the corresponding topic. Petitions over the period were related to a variety of topics. There were several petitions related to educational topics (i.e., petitions related to ragged schools), religious issues (i.e., petitions related to the Irish Church Bill) and various legal questions (i.e., petitions related to the property rights of married women or related to the sale of liquor). Examples are displayed in Appendix Figures A19A and A19B.

Information on petitions is available at the monthly level but not all months in our sample report petitions (i.e. our data is exhaustive but in some months no petitions were submitted). Overall, we have 3 months where we observe petitions before the elections and 7 months after the elections. The results are displayed in Table 4. To reflect the count data structure (see Appendix Figure A20), we rely on a Poisson model. The dependent variable of interest is the number of petitions received in month t from borough i. It is striking how an increase in the number of electors triggered a substantial boost in petitions, highlighting the multi-

dimensional surge in civic engagement in the aftermath of the Second Reform Act.[22] Finally, note that further results on mechanisms linked to voice are presented in Appendix E.2.

Table 4: Voice - Democracy and Petitions

Dep. Variable: Number of petitions submitted$_{it}$	(1)	(2)	(3)	(4)
(log) Electorate$_{it}$	3.274***	11.25***		
	(0.848)	(2.884)		
(log) Eligible Householders$_{it}$			1.805***	10.03***
			(0.531)	(1.439)
Observations	1,840	1,840	1,840	1,840
Borough FEs	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	No	Yes
Rent Inequality$_i$ ** Time FEs*	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers the period 1868-1869. Poisson estimates are reported in columns 1-4. The dependent variable is the number of petitions submitted in borough i and month t. The petition data was constructed using archival records of the Selected Committee on Public Petitions (see Parliament (n.d.)). The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.2 Increase in economic activity

A second channel that we investigate is that enfranchisement could have boosted economic activity consequently providing a fertile breeding ground for peace. In particular, more inclusive, pluralistic political institutions may create the conditions for more inclusive economic institutions and greater and more sustained economic growth (see Acemoglu et al., 2019 and Abeberesey et al., 2020, which suggest that democracy does favor economic growth). In turn, various papers (see e.g. Miguel et al., 2004; Dell et al., 2014; König et al., 2017) have found

[22] Given the scarcity of data points on petitions it is also useful to replicate this analysis for the extended sample. In particular, in Appendix Table A46 we report results with the extended sample (using the years 1866-1871) and including a linear trend (this corresponds to the specification of our extended sample specification of Table A25).

that peace is more easily achieved under favorable economic conditions.

Two factors that may make it more difficult in our context to detect large-scale effects on economic activity are that (i) our analysis is inherently short-run (and full-scale effects on economic activity may take time), and that (ii) we study a change in representation in the *national* parliament, while several –yet not all– relevant economic policy measures may be taken at the *local* level. Still, it is important to keep in mind that there could be anticipation effects (i.e. despite the short sample length, one may detect already a surge in economic activity early on if investment today is fuelled by the future prospects for peace and stability).

In our analysis, we use the number of times boroughs were mentioned in job advertisements in newspapers as a proxy of economic activity.[23] We show in Appendix B.4 that job advertisements correlate strongly with other proxies for economic activity from very different, existing sources.

The results are displayed in Table 5. Throughout all specifications, we find that greater enfranchisement was associated with an increase in our proxy of economic growth. The effect is of substantial magnitude: increasing the number of eligible electors by 89% (i.e. the average increase triggered by the reform) is associated with a rise in advertisements (our proxy for economic activity) by around 15.5%. Expressed in terms of standard deviations, a one standard deviation increase in enfranchisement is associated with greater economic activity of 0.16 standard deviations.

In the Appendix E.3 we further show that the strong association between the Second Act Reform and economic activity carries over to an alternative specification using a Poisson model (Appendix Table A47), and that our results are not driven by general media penetration (Appendix Table A48). We then demonstrate that the *predicted* level of job advertisements is indeed a strong determinant of lower social conflict levels (Appendix Table A49). Furthermore, we find that our results are robust to alternative definitions of our newspapers-based index of economic activity. These results are confined to Appendix E.4.

Finally, changes in economic activity do not only represent a channel of transmission through which enfranchisement reduces social conflict, but it is simultaneously an outcome of interest of its own, that may equally depend on the level of social conflict, i.e. there may also be an impact of protests (and the expectation of protests) on economic activity. Put differently, enfranchisement may at the same time boost economic growth and reduce conflict and these two beneficial effects may mutually reinforce each other (a better economic environment fosters peace and social stability fuels economic activity). The structure of our data would make it difficult to disentangle to what extent the impact of the economy on social peace is

[23] We identify pages of job advertisements as those pages classified as advertisement which contain the word "*wanted*".

larger than the inverse impact of social peace on economic activity.[24]

Table 5: Democracy and Economic Growth - Newspaper Ads

Dep. Variable: (log) Borough Mentions in Newsp. Ads$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	0.175***	0.167***	0.186***				0.248***	0.252***	0.241***
	(0.0446)	(0.0484)	(0.0458)				(0.0535)	(0.0604)	(0.0582)
(log) Eligible Householders$_{it}$				0.215***	0.226***	0.217***			
				(0.0481)	(0.0554)	(0.0526)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.952	0.952	0.954	0.952	0.952	0.954	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	3.974	3.974	3.974	3.974	3.974	3.974	3.974	3.974	3.974

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is (log+1) of number of mentions of borough i in pages of job advertisements containing the word "*wanted*" in month t using national or local newspapers available on the *British Newspaper Archive*. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.3 Increase in State Capacity

A third potential channel of transmission could be an increase in state capacity. As suggested by the 18th century political slogan "No taxation without representation", one may expect a quid-pro-quo with enfranchisement going along with an extension of the activities of the state. As argued, among others, by Fearon (2005); Collier, Hoeffler and Rohner (2009); Besley and Persson (2011*b*), weak state capacity can be a major cause for social conflict, hence a reason for the decline in social violence after franchise extension could be through greater state power.

Nevertheless, there are several reasons why it may be difficult to detect mechanisms linked to state capacity with our research design/data. First, one reason for scepticism on this potential channel is that it is well known among historians that this time period has not been characterized by high levels of public spending (see Aidt et al., 2010, Chapman et al., 2018, Chapman et al., 2020). Second, any potential proxies for state capacity for this period are

[24]For a theoretical framework featuring such virtuous cycles of business and peace mutually reinforcing each other, see e.g. Rohner, Thoenig and Zilibotti (2013).

quite unsatisfactory, which would typically result in substantial measurement error and attenuation bias. Third, our identification strategy focuses on the impact of *sharp* changes in the electorate on *sharp* short-run changes in state capacity. To the extent that building up powerful state infrastructure and enacting public spending takes many years, our identification strategy may miss out on such medium- and long-run effects. This is likely, for example, to apply to educational spending and school construction, which may deploy effects only several years down the road. Fourth, a factor that could lengthen delays in potential state capacity effects is the fact that this paper deals with elections for the national parliament and not for the local executive. While for the latter type of local executive elections any effects could be quite rapid and direct (i.e. a newly elected mayor may want to swiftly implement electoral promises), in our case of the former type of national legislative elections any effects would be indirect and could take longer (i.e. local constituents who are better represented at the national level are likely to benefit from more "pork-barrel" spending down the road some years later).[25] Altogether, these caveats imply that it would –if anything– be surprising to find state capacity related mechanisms over a short time horizon.
Nonetheless, in Appendix E.5 we study the impact of the Second Reform Act on (short-run) changes in public spending and deficits. As expected in the light of the aforementioned caveats, we are not able to detect such effects in the short-run.

6.4 Heterogeneous Effects

In the objective of further substantiating the notion that the Second Reform Act succeeded in attenuating social tensions by fostering economic opportunities, we study heterogeneous effects of the above results, by distinguishing between areas with big versus little economic potential. To be concise, all detailed information on the exact specification and all tables have been relegated to Appendix E.6. The most relevant results are that the pacifying and growth-promoting effects of enfranchisement are magnified in towns that have a high market potential (i.e. that are located close to large numbers of potential consumers). This is consistent with the notion that the growth-promoting effect of democratization may be one of the prime mechanisms at work that explains the drop in violence after the Second Reform Act.
A further heterogeneous effect makes use of detailed information on the demographic composition of the population. As discussed in depth in Appendix E.7, one dimension of social tensions in the 1860s were conflicts between the Anglican population and Catholic immigrant workers from Ireland, giving rise e.g. to the so-called "Murphy riots" (Arnstein, 1975). We ex-

[25] As discussed e.g. in Chapman et al. (2018) and Webster (2018), local councils were traditionally relatively important for public goods provision, but from the 1860s onwards the central government started to gain more and more importance in terms of financing. Importantly, Members of Parliament (MPs) played non-negligible roles for the financing of their local constituencies.

pect greater inter-group tensions in towns with a greater level of ethnic polarization (i.e. with a few large groups facing each other; e.g. in a borough with close to half of the population being English and the other half Irish). This is indeed what we detect in Appendix E.7 (which also contains all methodological details and exact variable definitions). These findings are in line with the notion that areas with higher initial social tensions experience a grater pacifying potential from political reform.

Finally, in Appendix E.8 we investigate what types of social violence are affected by enfranchisement. As explained in detail in Appendix E.8, we expect an (almost mechanical) decrease in *political* violence linked to claims for representation (as the enfranchisement has addressed various points of pre-reform demands), and also a reduction in *ethno-religious* violence due to better representation of all major ethnic groups in society. Finally, given the growth-promoting effect of the reform (see Section 6.2), one may expect a higher opportunity cost of social unrest, which could attenuate the risk of all types of social conflict – not only the those mentioned above, but also *economic* social conflicts. The results presented in Appendix E.8 show indeed that franchise extension tends to reduce all these three types of social conflict.

7. Conclusion

In this research paper, we examine the impact of enfranchisement on peace and prosperity, drawing on a milestone electoral reform of the UK's Victorian epoch: the Representation of the People Act of 1867. Building a novel panel dataset at the borough and month level for the period around the reform, and collecting novel data on social conflict, local economic growth and a battery of controls, this paper exploits arguably exogenous variation in the extent of enfranchisement across UK boroughs. Our results show a strong and significant pacifying effect of franchise extension that is reinforced by many successful robustness checks. While there is substantial evidence that a relevant channel of transmission is a surge in the population's political voice and competition in UK politics, we also find evidence for another mechanism at work which is that democratization fuels local economic growth.

On a more general level, our findings support the notion that civil peace and economic development are intertwined; it is difficult to cultivate peace in a context where economic growth is neglected, and vice-versa. This point of inter-dependence has been recently stressed by Rohner and Thoenig (2021) who talk about a *macro-complementarity* between promoting peace and fostering development. Our current paper highlights the complementary effect of a *political* reform not only achieving the *political* goal of reducing unrest, but on top of that boosting local *economic* growth. These complementarities of peace and prosperity call for an integrated approach of tackling unrest and under-development *at the same time*, rather than

leaving these inter-linked problems to two distinct policy communities. Further research on this, as well as an in-depth analysis of the effects of specific institutional rules, is strongly encouraged.

Bibliography

Abadie, A. and Gardeazabal, J. (2003), 'The economic costs of conflict: A case study of the Basque Country', *American Economic Review* **93**(1), 113–132.

Abeberesey, A. B., Barnwalz, P., Chaureyx, R. and Mukherjee, P. (2020), 'How does democracy cause growth? Evidence from firms during Indonesia's democratic transition'.

Acemoglu, D., Naidu, S., Restrepo, P. and Robinson, J. A. (2019), 'Democracy does cause growth', *Journal of Political Economy* **127**(1), 47–100.

Acemoglu, D. and Robinson, J. A. (2000), 'Why did the West extend the franchise? Democracy, inequality, and growth in historical perspective', *The Quarterly Journal of Economics* **115**(4), 1167–1199.

Acemoglu, D. and Robinson, J. A. (2001), 'A theory of political transitions', *American Economic Review* **91**(4), 938–963.

Acemoglu, D. and Robinson, J. A. (2005), *Economic origins of dictatorship and democracy*, Cambridge University Press.

Aidt, T. S., Daunton, M. and Dutta, J. (2010), 'The retrenchment hypothesis and the extension of the franchise in England and Wales', *The Economic Journal* **120**(547), 990–1020.

Aidt, T. S., Dutta, J. and Loukoianova, E. (2006), 'Democracy comes to Europe: Franchise extension and fiscal outcomes 1830–1938', *European Economic Review* **50**(2), 249–283.

Aidt, T. S. and Franck, R. (2015), 'Democratization under the threat of revolution: Evidence from the Great Reform Act of 1832', *Econometrica* **83**(2), 505–547.

Aidt, T. S. and Franck, R. (2019), 'What motivates an oligarchic elite to democratize? Evidence from the roll call vote on the Great Reform Act of 1832', *The Journal of Economic History* **79**(3), 773–825.

Aidt, T. S. and Jensen, P. S. (2014), 'Workers of the world, unite! Franchise extensions and the threat of revolution in Europe, 1820–1938', *European Economic Review* **72**, 52–75.

Anstey, T. C. (2022), *Notes Upon" The Representation of the People Act,'1867."*, BoD–Books on Demand.

Arnstein, W. L. (1975), 'The Murphy riots: a Victorian dilemma', *Victorian Studies* **19**(1), 51–

71.

Berlinski, S., Dewan, T. and Van Coppenolle, B. (2014), 'Franchise extension and the British aristocracy', *Legislative Studies Quarterly* **39**(4), 531–558.

Berlinski, S., Dewan, T. et al. (2011), 'The political consequences of franchise extension: Evidence from the Second Reform Act', *Quarterly Journal of Political Science* **6**(34), 329–376.

Besley, T. and Kudamatsu, M. (2006), 'Health and democracy', *American Economic Review* **96**(2), 313–318.

Besley, T. and Persson, T. (2011*a*), 'The logic of political violence', *The Quarterly Journal of Economics* **126**(3), 1411–1445.

Besley, T. and Persson, T. (2011*b*), *Pillars of prosperity: Te political economics of development clusters*, Princeton University Press.

Bidner, C., Francois, P. and Trebbi, F. (2014), 'A theory of minimalist democracy', *NBER Working Paper* (w20552).

Billmeier, A. and Nannicini, T. (2013), 'Assessing economic liberalization episodes: A synthetic control approach', *Review of Economics and Statistics* **95**(3), 983–1001.

Caprettini, B. and Voth, H.-J. (2020), 'Rage against the machines: Labor-saving technology and unrest in industrializing england', *American Economic Review: Insights* **2**(3), 305–20.

Cederman, L.-E., Gleditsch, K. S. and Hug, S. (2013), 'Elections and ethnic civil war', *Comparative Political Studies* **46**(3), 387–417.

Cervellati, M. and Sunde, U. (2013), 'Democratizing for peace? The effect of democratization on civil conflicts', *Oxford Economic Papers* **66**(3), 774–797.

Chapman, J. et al. (2018), 'Democratic reform and opposition to government expenditure: Evidence from nineteenth-century Britain', *Quarterly Journal of Political Science* **13**(4), 363–404.

Chapman, J. et al. (2020), 'Democracy, redistribution, and inequality: Evidence from the English Poor Law', *Working Paper* .

Cilliers, J., Dube, O. and Siddiqi, B. (2016), 'Reconciling after civil conflict increases social capital but decreases individual well-being', *Science* **352**(6287), 787–794.

Colella, F., Lalive, R., Sakalli, S. O. and Thoenig, M. (2019), 'Inference with arbitrary clustering', *IZA Discussion Paper* .

Collier, P., Hoeffler, A. and Rohner, D. (2009), 'Beyond greed and grievance: Feasibility and civil war', *Oxford Economic Papers* **61**(1), 1–27.

Collier, P. and Rohner, D. (2008), 'Democracy, development, and conflict', *Journal of the European Economic Association* **6**(2-3), 531–540.

Collier, P. and Vicente, P. C. (2014), 'Votes and violence: Evidence from a field experiment in Nigeria', *The Economic Journal* **124**(574), F327–F355.

Conley, T. G. (1999), 'Gmm estimation with cross sectional dependence', *Journal of Econometrics* **92**(1), 1–45.

Craig, F. W. S. (1977), *British parliamentary election results*, Springer.

Davis, J. and Tanner, D. (1996), 'The borough franchise after 1867', *Historical Research* **69**(170), 306–327.

Dell, M., Jones, B. F. and Olken, B. A. (2014), 'What do we learn from the weather? The new climate-economy literature', *Journal of Economic Literature* **52**(3), 740–98.

Dercon, S. and Gutiérrez-Romero, R. (2012), 'Triggers and characteristics of the 2007 Kenyan electoral violence', *World Development* **40**(4), 731–744.

Doepke, M. and Tertilt, M. (2009), 'Women's liberation: What's in it for men?', *The Quarterly Journal of Economics* **124**(4), 1541–1591.

Durante, R. and Zhuravskaya, E. (2018), 'Attack when the world is not watching? us news and the israeli-palestinian conflict', *Journal of Political Economy* **126**(3), 1085–1133.

Esteban, J., Mayoral, L. and Ray, D. (2012), 'Ethnicity and conflict: An empirical study', *American Economic Review* **102**(4), 1310–42.

Esteban, J., Morelli, M. and Rohner, D. (2015), 'Strategic mass killings', *Journal of Political Economy* **123**(5), 1087–1132.

Facchini, G., Knight, B. G. and Testa, C. (2020), 'The franchise, policing, and race: Evidence from arrests data and the Voting Rights Act'.

Fearon, J. D. (2005), 'Primary commodity exports and civil war', *Journal of Conflict Resolution* **49**(4), 483–507.

Fearon, J. D. (2011), 'Self-enforcing democracy', *The Quarterly Journal of Economics* **126**(4), 1661–1708.

Fearon, J. D. and Laitin, D. D. (2003), 'Ethnicity, insurgency, and civil war', *American Political Science Review* **97**(1), 75–90.

Fergusson, L., Querubin, P., Ruiz, N. A. and Vargas, J. F. (2020), 'The real winner's curse', *American Journal of Political Science* .

Fergusson, L., Riano, J. F. and Song, B. (2020), 'Media, secret ballot and democratization in the US'.

Fetzer, T. (2020), 'Can workfare programs moderate conflict? evidence from india', *Journal of the European Economic Association* **18**(6), 3337–3375.

Fetzer, T. and Kyburz, S. (2018), 'Cohesive institutions and political violence', *The Review of Economics and Statistics* pp. 1–46.

Fraser, D. (1976), *Urban politics in modern England*, Springer.

Fujiwara, T. (2015), 'Voting technology, political responsiveness, and infant health: Evidence from brazil', *Econometrica* **83**(2), 423–464.

Harris, C. D. (1954), 'The market as a factor in the localization of industry in the United States', *Annals of the Association of American Geographers* **44**(4), 315–348.

Harrison, R. (1962), 'The 10th april of spencer walpole: The problem of revolution in relation to reform, 1865–18671', *International Review of Social History* **7**(3), 351–399.

Hegre, H., Ellingsen, T., Gates, S. and Gleditsch, N. P. (2001), 'Toward a democratic civil peace? Democracy, political change, and civil war, 1816-1992', *American Political Science Review* **95**(1), 33–48.

Himmelfarb, G. (1966), 'The politics of democracy: the English Reform Act of 1867', *The Journal of British Studies* **6**(1), 97–138.

Holland, M. (2005), *Swing unmasked: the agricultural riots of 1830 to 1832 and their wider implications*, FACHRS Publishing.

House of Commons (1857), *Accounts and Papers of the House of Commons, Vol. XXXIV - Sess. 30 April – 28 August 1857*.

House of Commons (1860*a*), *Accounts and Papers of the House of Commons, Vol. LV. - Sess. 24 January – 28 August 1860*.

House of Commons (1860*b*), *Parliamentary Cities and Boroughs - Return to an Order of the Honourable the House of Commons dated 17 May 1860*.

House of Commons (1866*a*), *Accounts and Papers of the House of Commons, Vol. LVII. - Sess. 1 February– 10 August 1866*.

House of Commons (1866*b*), *Electoral Statistics - Return to an Address of the Honourable the*

House of Commons dated 7 August 1866.

House of Commons (1869*a*), *Electoral Statistics - Return to an Address of the Honourable the House of Commons dated 7 August 1866*.

House of Commons (1869*b*), *Electors in Cities and Boroughs - Return to an Address of the Honourable the House of Commons dated 12 March 1869*.

House of Commons (1873), *Income Tax - Parliamentary Constituencies - Return to an Order of the Honourable the House of Commons dated 1 July 1873*.

House of Commons (1874*a*), *Accounts and Papers of the House of Commons, Vol. LIII. - Sess. 5 March – 7 August 1874*.

House of Commons (1874*b*), *Electoral Statistics - Return to an Address of the Honourable the House of Commons dated 1 May 1874*.

House of Commons (1877), *Electoral Statistics - Return to an Address of the Honourable the House of Commons dated 11 May 1877*.

IPUMS (2020), '1861 Population Census - Minnesota population center. integrated public use microdata series, international: Version 7.2', https://international.ipums.org/international/.

Jetter, M. (2017), 'The effect of media attention on terrorism', *Journal of Public Economics* **153**, 32–48.

Knatchbull-Hugessen, E. H. (1869), *Abstract of the Statements of Accounts of the several Municipal Boroughs in England and Wales for the Year ending 31st August 1868*.

Knatchbull-Hugessen, E. H. (1870), *Abstract of the Statements of all monies Received and Expended on Account of the Mayor, Aldermen, and Burgesses of each of the different Municipal Boroughs, for the Year ending 31st August 1869*.

König, M. D., Rohner, D., Thoenig, M. and Zilibotti, F. (2017), 'Networks in conflict: Theory and evidence from the Great War of Africa', *Econometrica* **85**(4), 1093–1132.

Kudamatsu, M. (2012), 'Has democratization reduced infant mortality in sub-saharan Africa? Evidence from micro data', *Journal of the European Economic Association* **10**(6), 1294–1317.

Lacroix, J. (2020), 'Ballots instead of bullets? The effect of the Voting Rights Act on political violence'.

Laurent-Lucchetti, J., Rohner, D. and Thoenig, M. (2019), 'Ethnic conflicts and the informational dividend of democracy'.

Lawrence, J. (2009), *Electing our masters: the hustings in British politics from Hogarth to Blair*, Oxford University Press.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. and Zettlemoyer, L. (2019), 'Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension', *arXiv preprint arXiv:1910.13461* .

Lizzeri, A. and Persico, N. (2004), 'Why did the elites extend the suffrage? Democracy and the scope of government, with an application to Britain's "Age of Reform"', *The Quarterly Journal of Economics* **119**(2), 707–765.

Llavador, H. and Oxoby, R. J. (2005), 'Partisan competition, growth, and the franchise', *The Quarterly Journal of Economics* **120**(3), 1155–1189.

Mann, M. (2005), *The dark side of democracy: Explaining ethnic cleansing*, Cambridge University Press.

Miguel, E., Satyanath, S. and Sergenti, E. (2004), 'Economic shocks and civil conflict: An instrumental variables approach', *Journal of Political Economy* **112**(4), 725–753.

Montalvo, J. G. and Reynal-Querol, M. (2005), 'Ethnic polarization, potential conflict, and civil wars', *American Economic Review* **95**(3), 796–816.

Mueller, H. and Rohner, D. (2018), 'Can power-sharing foster peace? Evidence from Northern Ireland', *Economic Policy* **33**(95), 447–484.

Nunn, N. and Qian, N. (2014), 'US food aid and civil conflict', *American Economic Review* **104**(6), 1630–66.

O'Brien, P. K. (1959), 'British incomes and property in the early nineteenth century', *The Economic History Review* **12**(2), 255–267.

Olea, J. L. M. and Pflueger, C. (2013), 'A robust test for weak instruments', *Journal of Business & Economic Statistics* **31**(3), 358–369.

Parliament, G. B. (n.d.), *Public Petitions to Parliament, 1833-1918*, Proquest.

Przeworski, A. (2009), 'Conquered or granted? A history of suffrage extensions', *British Journal of Political Science* **39**(2), 291–321.

Rohner, D. (2022), 'Mediation, military and money: The promises and pitfalls of outside interventions to end armed conflicts', *Journal of Economic Literature* .

Rohner, D. and Saia, A. (2019), 'Education and conflict: Evidence from a policy experiment in Indonesia'.

Rohner, D. and Thoenig, M. (2021), 'The elusive peace dividend of development policy: From war traps to macro-complementarities', *Annual Review of Economics* **13**, 111–131.

Rohner, D., Thoenig, M. and Zilibotti, F. (2013), 'War signals: A theory of trade, trust, and conflict', *Review of Economic Studies* **80**(3), 1114–1147.

Saia, A. (2017), 'Choosing the open sea: The cost to the UK of staying out of the euro', *Journal of International Economics* **108**, 82–98.

Saunders, R. (2007), 'The politics of reform and the making of the Second Reform Act, 1848-1867', *Historical Journal* pp. 571–591.

Schlager, N. (2004), *St. James Encyclopedia of Labor History Worldwide: Major Events in Labor History and Their Impact*, Vol. 2, Saint James Press.

Smellie, K. B. (1968), *A history of local government*, number 1, Allen & Unwin.

Snyder, J. L. (2000), *From voting to violence: Democratization and nationalist conflict*, Norton New York.

Turner, J. D. and Zhan, W. (2012), 'Property rights and competing for the affections of Demos: The impact of the 1867 Reform Act on stock prices', *Public Choice* **150**(3-4), 609–631.

Von Hippel, P. T., Hunter, D. J. and Drown, M. (2017), 'Better estimates from binned income data: Interpolated CDFs and mean-matching', *Sociological Science* **4**, 641–655.

von Hippel, P. T., Scarpino, S. V. and Holas, I. (2016), 'Robust estimation of inequality from binned incomes', *Sociological Methodology* **46**(1), 212–251.

Webster, I. (2018), 'The public works loan board and the growth of the state in nineteenth-century england', *The Economic History Review* **71**(3), 887–908.

Yin, W., Hay, J. and Roth, D. (2019), 'Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach', *arXiv preprint arXiv:1909.00161* .

Zimmerman, K. (2003), 'Liberal speech, Palmerstonian delay, and the passage of the Second Reform Act', *The English Historical Review* **118**(479), 1176–1207.

Appendix

In the Online Appendices below we provide additional description, investigation and further results for the various sections of this paper. We first provide additional information on the historical context in Appendix Section A, focusing on enfranchisement rules (Appendix A.1) and political incentives for the reform (Appendix A.2).
Next, in Appendix Section B, we provide a series of descriptive summary statistics (Appendix B.1) and describe in more depth the data on rents (Appendix B.2), social conflict (Appendix B.3) and job advertisements (Appendix B.4). We also provide "validation" of the social conflict and job ads data, drawing on existing, external data sources.
Moreover, in Section C we display the raw association between the instrument and change in social conflict.
Furthermore, the voluminous Appendix Section D presents a wide array of robustness checks, focusing on additional controls (Appendices D.1-D.5), alternative instruments (Appendices D.6-D.7), alternative dependent variables (Appendix D.8), estimation methods and inference (Appendices D.9-D.12), outliers and different time units (Appendices D.13-D.14), time trends and extended sample period (Appendix D.15), placebo analysis (Appendices D.16-D.19) and data construction (Appendices D.20-D.24).
The following Appendix Section E provides additional information and results on potential channels and mechanisms, namely on increased voice (Appendices E.1-E.2), economic activity (Appendices E.3-E.4), and state capacity (Appendix E.5). We also report the results of the analysis of heterogenous effects related to market potential (Appendix E.6), polarization (Appendix E.7) and different event types (Appendix E.8).
Lastly, we list supplementary data on newspapers used and borough-level rent statistics in the Appendix Sections F and G, respectively.
Below is listed the Table of Content of the Online Appendices.

A. Appendix: Historical Context

A.1 Additional Details on Enfranchisement Rules

In what follows we provide additional details on particular aspects of the historical context, notably on the specific enfranchisement rules, and discuss how we take this into account in the variable construction.

Prior to the 1867 reform bill, males living in urban areas were allowed to vote if they were owners or occupiers of a property with an annual rental value of at least £10 (i.e., the estimated cost of renting the property for a year (Saunders, 2007)), they were paying taxes and they had resided in the property for at least one year. For owner occupiers the rental value of the house was imputed on the basis of the house characteristics without taking into account the furniture.

The 1867 Reform Act extended voting rights to all householders (owners or occupiers) of houses in English and Welsh boroughs that had resided during the last 12 months and paid their taxes (i.e. the so-called "poor rates").[26][27]

The reform extended voting rights also to males in rural areas but the impact was limited. The 1867 reform bill extended the franchise living in rural areas "*to those who owned or had leases of 60 years or longer on land valued at £5 or more per annum and occupiers of land which had a rateable value of £12 or more and who had paid the appropriate poor-law rates*" (Turner and Zhan, 2012). This being said, we focus in our analysis on the sample of boroughs, which are more urban than the counties, because i) the boroughs experienced a larger increase in terms of voters, ii) data for urban areas (both in terms of outcomes and control variables) are generally more available and of higher quality, iii) our instrument only applies to boroughs (i.e., the previous administrative threshold of £10 rental value was not in place in rural areas) and iv) counties contain dozens of small parishes, making the geolocation of a violent event potentially less precise.

An important aspect is that –as discussed above– householders were only enfranchised when

[26] *The individual must have paid during the time of such occupation "all rates (if any) made for the relief of the poor"*, see Anstey (2022), page 151.

[27] Note that the Reform Act distinguished between tenants who occupy a whole building ("householders") versus renters of a single room ("lodgers"). The Reform Act enfranchised also lodgers with a £10 annual rental value. Overall, the number of electors entitled as lodgers on Register of 1869 is fairly low compared to the voters entitled as occupiers. Overall, the occupiers represents the overwhelming majority of individuals entitled to vote (for example, in Liverpool the number of occupiers and the number of lodgers on the Register of 1869 were 38,000 and 529, respectively (on the Register of 1866 they were 20,554 and 0), in Manchester 48,228 and 28 (in 1866 they were 22,792 and 0), in Birmingham 42,041 and 1 (in 1866 they were 15,490 and 0), in Cardiff 5,099 and 61 (in 1866 they were 1,888 and 0)). A notable exception was the borough of Westminster were the increase in the electors observed in 1869 was mostly driven by the lodger group. In Westminster the number of occupiers and lodgers observed in the Register of 1869 was 14,348 and 4,307 (the number recorded in 1866 was 12,215 and 0, respectively) (source: House of Commons (1874*b*)).

they pay taxes (i.e. the so-called "poor rates"). Only citizens with some minimum level of income were subject to taxation, and a key question is how this threshold translates into the average house value of a given householder. In our baseline analysis we set this threshold at £4 annual rental value, hence presuming that citizens living in below-threshold housing typically did not pay the poor rates taxes (and hence were not enfranchised) while those living in houses with value above the threshold did pay taxes (and were enfranchised by the Second Reform Act). Importantly, in robustness checks below we show that our results are not sensitive to the exact threshold level.

Having as threshold the £4 rental value is reasonable. The Law Times [July 4, 1868][28] provided a table showing the additions to be made to the numbers of the borough electors by the new householder franchise: "*We have taken the trouble to analyse the poor-rate returns of all the boroughs, for the purpose of ascertaining the probable numbers which household suffrage will add to each of them.*" While setting a threshold of rental value at 0 pounds would lead to overstating the number of new electors by almost half, while setting a threshold of 4 pounds leads to a number of estimated new voters very close to the actual numbers.[29]

This is also consistent with a Return to an Address of the Honourable The House of Commons, dated the 17 May 1860, which showed for every parliamentary borough and borough in England and Wales the number of persons rated as occupiers to the Relief of the Poor in 1853, and where the lowest threshold displayed in the table was £4. This clearly shows that below 4 pounds citizens were very unlikely to pay taxes.

Finally, in line with this, the Report of the select committee appointed to inquire what would be the probable increase of the number of electors in the counties and boroughs of England and Wales from a reduction of the franchise [1860] stated: "*Question - Then, in order to make this return accurate, as an estimate of the effect of the proposed change, it is necessary to add all the occupiers of houses put down in your rate book at a gross estimated rented of £6 or £4?*" "*Answer - Yes, £4 and above.*" [House of Commons (1860*a*), Page 53].

A.2 Who Benefitted From the Reform?

One relevant question to ask is whether the incumbent Conservatives led by Prime Minister Benjamin Disraeli could benefit electorally from the reform and may have been motivated politically to promote the reform act. A priori, this is not very likely, as (i) the reform act was a bi-partisan bill benefiting also from the support of the opposition, and (ii) in the 1868

[28] The Law Times was a periodical, published from 1843 to 1965. It contained, among other things, information on all the cases treated and decided in the House of Lords.

[29] The average percentage difference between the full set of householders below £10 and the probable addition of electors reported in the Law Times is of around 42%, whereas the average percentage difference between the full set of householders above £4 and below £10 and the probable addition of electors reported in the Law Times is only of around 8%.

election the incumbent Conservatives were defeated by the Liberals, paving the way for the Liberal William Gladstone to become the new Prime Minister.

To go one step further, we display below Table A1 that regresses the local electoral support for Conservative candidates on the extent of local enfranchisement. We are not able to detect any robust and significant link between enfranchisement and the relative propensity to vote for the Conservatives. Our findings are in line with those of Berlinski et al. (2011) who also do not detect a change in relative party support in the aftermath of the reform act.

Table A1: The impact of enfranchisement on voting outcomes in the 1868 election

	(1)	(2)	(3)	(4)
Dep. Variable:	*Votes Share Cons. Party*$_{it}$	*Share Seats Cons. Party*$_{it}$	*Elected Num. Cons. Party*$_{it}$	*Share Seats Inc. Cons. Party*$_{it}$
(log) Electorate$_{it}$	-0.0124	0.0290	0.0634	-0.0717
	(0.0509)	(0.0808)	(0.134)	(0.0614)
Observations	366	366	366	366
R-squared	0.801	0.712	0.738	0.830
Borough FEs	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	Yes	Yes	Yes	Yes
Rent Inequality$_i$ ** Time FEs*	Yes	Yes	Yes	Yes
Seats	Yes	No	Yes	No

Notes: The unit of observation is a borough i and election year t. The sample covers 183 boroughs over the elections in years 1865 and 1868. OLS estimates are reported in columns 1-4. In each column we display results for different dependent variables relative to the conservative party. In column (1) the dependent variable is the vote share for the conservative party in each election (i.e. 1865-1868), in column (2) it is the share of elected conservative candidates over the number of seats, in column (3) it is the number of elected conservative candidates and in column (4) it is the share of incumbent conservative candidates that get re-elected. These dependent variables were constructed using data collected by the authors from Craig (1977). The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B. Appendix: Data

B.1 Sample Composition and Overall Summary Descriptive Statistics

This appendix section provides an illustration and further information on the data included. In Table A2 the summary descriptive statistics are displayed. As discussed in the main text above in Section 3.5, about 23 percent of observations experienced social conflict and the period of interest in the current study (1866 to 1869) displayed a much larger increase in electors than any of the other periods before or after.

Table A2: Descriptive Statistics

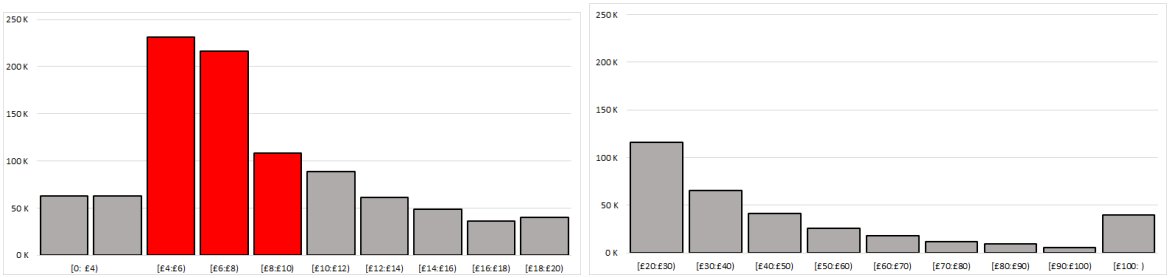
Variable	Mean	Std. Dev.	Min.	Max.	N
Social Violence$_{it}$	.237	0.425	0	1	4,416
(log) Electorate$_{it}$	7.618	1.193	5.165	10.78	4,416
Δ *Electorate 1873 - 1869*$_i$	0.071	0.291	-2.61	1.062	180
Δ *Electorate 1869 - 1866*$_i$	0.892	0.403	-0.047	2.352	184
Δ *Electorate 1866 - 1860*$_i$	0.092	0.135	-0.379	0.570	184
Δ *Electorate 1860 - 1857*$_i$	0.029	0.081	-0.257	0.276	182
Δ *Electorate 1857 - 1832*$_i$	0.267	0.445	-0.822	2.988	182
(log) Eligible Householders$_{it}$	7.604	1.239	5.247	11.390	4,416
Δ *Eligible Householders in 1866*$_i$	0.823	0.354	0.01	1.899	184
Average Rents$_i$	26.587	17.262	1.445	184.89	184
Rent Inequality$_i$	0.494	0.071	0.238	0.641	184
(log) *Voters*$_{it}$	7.454	1.156	5.247	10.523	216
Δ *Voters 1868-1865*$_i$	0.788	0.524	-1.926	1.819	108
(log) Borough Mentions in Newspaper Ads +1$_{it}$	3.974	1.323	0	7.472	4,416
Number of petitions submitted$_{it}$	2.609	7.292	0	100	1,840

NOTE: For sources and details see Section 3.

B.2 Data on Rents

As discussed above, the entry-threshold of four pounds for belonging to the category of new voters is somewhat less clear-cut and was less explicitly communicated to the population at large (which is why we perform a battery of robustness checks below, showing that our results continue to hold in the absence of this lower threshold). In contrast, the exit-threshold of £10 (above which one could already vote before the reform act) was widely known at the time and heatedly discussed. Hence, if there were to be any risk of strategic manipulation of rents, it would be around the ten pound threshold. To investigate this possibility, we display graphically the distribution of rents in Figure A1 below. Importantly, we do not detect any indication of a jump between the rent categories of "between 8 and 10" and "between 10 and 12", respectively, and hence there does not appear to be a plausible risk of strategic rent manipulation. To further address such concerns, below in the robustness analysis we show that our results also go through when using rents data from more than ten years earlier (1853).

Figure A1: Rent Bands Distribution



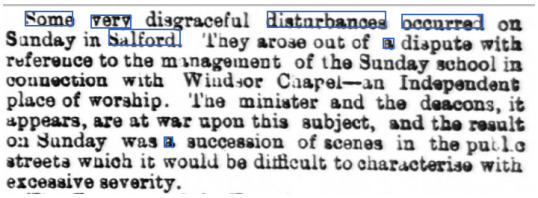
(A) Below £20 (B) Above £20

NOTE: Rents distribution summing the number of individuals over the 184 boroughs present in the main sample. Value corresponds to the number of male persons occupiers at different rental values (based on the estimated cost of renting the property for a year). In red are displayed the rent bands which correspond to the householders enfranchised by the reform. For consistency reasons, total number of householders in the rent band [0:£4) are splitted over two columns. The rent aggregation corresponds to the one used in Columns (1) and (2) of Table A9.

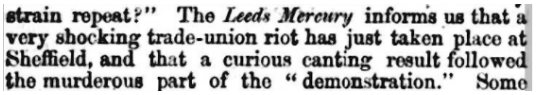
B.3 Data on Social Conflict

In this subsection we discuss in more detail the data generation and provide examples of social conflict events. In a second step we "validate" the social conflict data generated by our algorithm by comparing them with an independent historical dataset on the "Swing Riots". First of all, for the sake of illustration, Figure A2 displays four examples from media reports on typical social conflict events that we pick up with our algorithm.

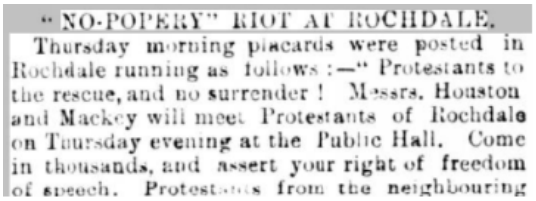
Figure A2: Conflict Examples


Some very disgraceful disturbances occurred on Sunday in Salford. They arose out of a dispute with reference to the management of the Sunday school in connection with Windsor Chapel—an Independent place of worship. The minister and the deacons, it appears, are at war upon this subject, and the result on Sunday was a succession of scenes in the public streets which it would be difficult to characterise with excessive severity.

(A) Article from *Blackburn Standard* of Wednesday 19 August 1868

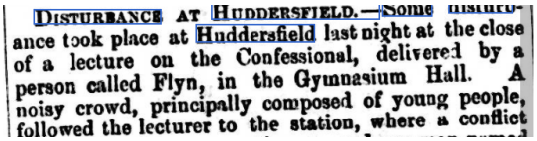

strain repeat?" The *Leeds Mercury* informs us that a very shocking trade-union riot has just taken place at Sheffield, and that a curious canting result followed the murderous part of the "demonstration." Some

(B) Article from *The Sportsman* of Thursday 19 August 1869


"NO-POPERY" RIOT AT ROCHDALE.

Thursday morning placards were posted in Rochdale running as follows:—"Protestants to the rescue, and no surrender! Messrs. Houston and Mackey will meet Protestants of Rochdale on Thursday evening at the Public Hall. Come in thousands, and assert your right of freedom of speech. Protestants from the neighbouring

(C) Article from *Dublin Evening Post* of Saturday 07 March 1868


DISTURBANCE AT HUDDERSFIELD.—Some disturbance took place at Huddersfield last night at the close of a lecture on the Confessional, delivered by a person called Flyn, in the Gymnasium Hall. A noisy crowd, principally composed of young people, followed the lecturer to the station, where a conflict

(D) Article from *Globe* of Saturday 27 June 1868

Further, we now confront our data collection from historical newspaper articles to existing data from an external source. While we were not able to retrieve historical borough or county level data for the period of study, we have been able to validate our method to extract social conflict event data from historic newspaper articles for an earlier period where social violence-related data is available. To do so, we use data from the *Swing Riots* from Holland (2005). While the original dataset contains all Captain Swing riots that occurred over the period January 1830 to December 1832, we follow Caprettini and Voth (2020), and only consider episodes that occurred between August 1830 and December 1832. This enables us to leverage on the geo-referencing done by Caprettini and Voth (2021) to link the location of an event to the corresponding county.

As a next step, we have followed the procedure outlined in the main text to construct social violence data using articles from the British Newspaper Archive. In particular, we have downloaded all articles for the period August 1830 until December 1832 containing violence-related keywords. To identify social conflict-related events, we have analyzed the content of all articles and we have checked if the name of a parish is located in the string [-25:+50]

nearby social conflict-related words. Finally, we have merged all corresponding events to the corresponding county.

After the steps outlined above, we have two sources of conflict data available, namely the external Swing Riots data, as well as our social conflict data from newspaper articles available on the British Newspaper Archive, for a total of 1,595 observations (55 counties over a period of 29 months from August 1830 until December 1832).

A few caveats are worthy of discussion before presenting the results. Firstly, the precision of the geolocation of an event is likely to be reduced when we consider names of parishes. For example, if a riot occurs in a small parish, it is likely that the article uses the name of the county or the area to help the reader to identify the location of the event (which is typically less of a problem when large boroughs are considered, as in our main analysis). Secondly, the quality of OCR (optical character recognition) decreases when we consider articles from an earlier period (i.e. our dataset for the main analysis is over thirty years more recent than the Swing Riot data for the validation exercise). Finally, with our methods we are not able to identify the exact nature or cause of the events. As a consequence, our social conflict variable includes all forms of social violence (including but not limited to the Swing riot episodes). Taken together, these caveats imply that we expect a correlation that is positive but not necessarily very large.

Table A3 below reports the correlations between swing riots data and our social conflict data over the period August 1830 to December 1832. The dependent variable is the total number of swing riots observed in the corresponding county-month. In column 1 only the social conflict measures obtained from historical newspaper data are used as explanatory variables. The coefficients of interest indicate a positive and statistically significant association between our measure and the swing riots data. The association remains positive and statistically significant when we include both month fixed effects (column 2) as well as county fixed effects (column 3). Overall, the results displayed in Table A3 provide support for our method of retrieving social conflict from historical newspaper articles.

Finally, we present a set of robustness exercises to further validate our measure. Table A4 presents results of specifications that include time-varying effects of population in the county. Table A5 restricts the sample to the county-month observations that experienced at least one Swing riot episode.

Below in Appendix Section D, we provide further support for the validity of our data by conducting a battery of robustness tests that include but are not limited to i) manual checking of sentences, ii) alternative algorithms and keywords to identify events, iii) applying machine learning tools to predicting true positive events, and iv) using Natural Language Processing (NLP).

Table A3: Swing Riots and social conflict data using articles from the British Newspaper Archive - County-month level

Dep. Variable: Swing Riots$_{it}$	(1)	(2)	(3)
Social Violence from Newspapers$_{it}$	1.032***	0.934***	0.936***
	(0.154)	(0.155)	(0.178)
Observations	1,595	1,595	1,595
R-squared	0.179	0.308	0.356
Month FEs	No	Yes	Yes
County FEs	No	No	Yes

Notes: The unit of observation is county i and month t. The sample covers 55 counties over a period of 29 months from August 1830 until December 1832. OLS estimates are reported in all columns. The dependent variable is the number of Swing riots events that were observed in county i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *Social Conflict*$_{it}$ corresponds to the number of events observed in county i and month t. Robust standard errors clustered at the county level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Swing Riots and social conflict data using articles from the British Newspaper Archive - County-month level - Flexible controls for population

Dep. Variable: Swing Riots$_{it}$	(1)	(2)	(3)
Social Violence from Newspapers$_{it}$	1.000***	1.000***	0.933***
	(0.166)	(0.166)	(0.186)
Observations	1,595	1,595	1,595
R-squared	0.328	0.328	0.372
Month FEs	No	Yes	Yes
County FEs	No	No	Yes

Notes:The unit of observation is county i and month t. The sample covers 55 counties over a period of 29 months from August 1830 until December 1832. OLS estimates are reported in all columns. The dependent variable is the number of Swing riots events that were observed in county i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *Social Conflict*$_{it}$ corresponds to the number of events observed in county i and month t. All regressions include the interaction between total population in the county (computed as the total population of all parishes in the county using data from Caprettini and Voth (2020)) and month-year fixed effects. Robust standard errors clustered at the county level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Swing Riots and social conflict data using articles from the British Newspaper Archive - County-month level - Only County-Month with at least one Swing Riot event

Dep. Variable: Swing Riots$_{it}$	(1)	(2)	(3)
Social Violence from Newspapers$_{it}$	1.322***	1.301***	1.070***
	(0.180)	(0.196)	(0.242)
Observations	449	449	446
R-squared	0.178	0.352	0.413
Month FEs	No	Yes	Yes
County FEs	No	No	Yes

Notes:The unit of observation is county i and month t. The full sample covers 55 counties over a period of 29 months from August 1830 until December 1832. The table focuses on the county-month observations where at least one swing riot event was observed. OLS estimates are reported in all columns. The dependent variable is the number of Swing riots events that were observed in county i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *Social Conflict*$_{it}$ corresponds to the number of events observed in county i and month t. Robust standard errors clustered at the county level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.4 Data on Job Ads

A key variable of the analysis of mechanisms and channels of transmission is our novel fine-grained data on job ads, which has been generated from newspaper articles and which we describe in some depth in the main text. As this variable is used as proxy for economic performance, it is useful to compare it to existing administrative data. Below, we perform several validation exercises for the job ads variable. Note that in addition to the exercises presented in the current Appendix Section (which focus on validation using existing administrative data), we have also carried out a series of sensitivity and robustness checks, where alternative ways of constructing the job ads variable and placebo exercises have been performed (see Section E.3).

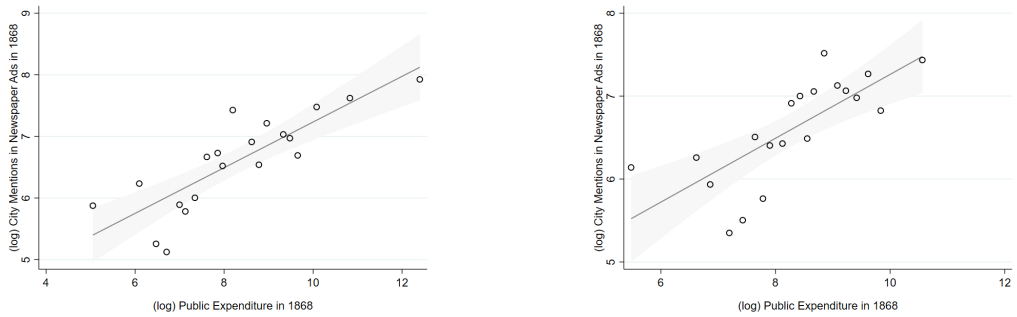
Job Ads Variable – Validation with Public Expenditure Data Here we compare the job ads variable with existing administrative data on public expenditures, which is however only available at the borough level for 1868 (and only at the annual, not monthly level). Incidentally, this is a key reason why we cannot directly use public expenditures as a proxy for measuring the economic effects of the reform. However, while we are limited in the use of public expenditures as an outcome variable in our main regressions, we can correlate it in the cross-section with job ads. As public expenditures are arguably a powerful proxy for economic development, a strongly positive correlation of public expenditures and job ads would reassure us on the reliability of job ads as proxy for economic development. In Figure A3 below we display the correlation between borough-level annual job ads in 1868 and borough-level public expenditures in the same year. The left panel of the figure displays the raw correlation, while the right panel of the figure displays the conditional correlation controlling for the population in the borough. The results highlight a strong positive correlation of our measure of economic activity based on newspaper ads and public economic activity (measured using administrative public expenditure data).

Job Ads Variable – Validation with Tax Data Another useful proxy for economic development that we were able to retrieve from the British historical archives is the *Total Number of Persons paying income taxes under Schedule D* over the period 1870-1872 at the borough level (House of Commons, 1873).[30] As this data is only available for the post-reform period, we cannot use it as a dependent variable in the main regressions, but we can –analogous to the exercise above– use this data to validate our job ads variable.

We carry out two validation exercises. First, we display the cross-sectional correlation between our job ads measure and the number of individuals paying taxes under Schedule D,

[30] As pointed out by O'Brien (1959), income taxes under schedule D can be "regarded as excluding wages and consisting mainly of industrial and commercial profits, but including, in addition, the salaries of what today are called the professional and managerial classes" (1959: 260). While we may have preferred a more coherent tax base, this variable still well serves the role of a proxy for economic development.

Figure A3: Public Expenditure and Job Ads



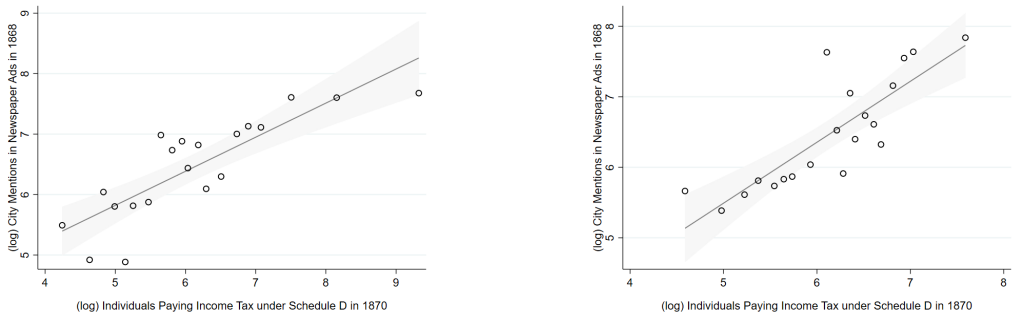
(A) Unconstrained (B) Controlling for Population

NOTE: Bin scatter graph with each bin summarizing values of several boroughs. The red line represents the linear prediction plot based on the underlying data. Panel A displays the values of (log) borough mentions in newspaper ads and the (log) Public Expenditure in 1868 along two axes. Panel B displays the values of the residuals of the values of (log) borough mentions in newspaper ads and the (log) Public Expenditure in 1868 along two axes when controlling for population in the borough along two axes.

using information on the individuals who were paying taxes (under schedule D) at the borough level observed in 1870. Figure A4 depicts the correlation between job ads at the yearly level in 1870 and the tax-related information in the same year. The left panel reports the raw correlation, while the right panel shows the conditional correlation controlling for the population in the borough. We find a strong correlation between our measure of economic activity measured using newspaper ads and economic activity proxied by the number of individuals subject to taxation.

Second, we exploit the time series dimension (1870-1872) to highlight that the correlation between job ads and taxes continues to hold even in a panel when controlling for borough fixed effects and year fixed effects. In particular, Table A6 reports the correlations between the income tax data and our ads data for the 1870-1872 period. The dependent variable used in columns 1-3 corresponds to the (log) total number of individuals paying taxes under Section D in the corresponding borough-year. In column 1 only the (log) total number of ads observed in the corresponding year is used as explanatory variable. The coefficient of interest indicates a positive and statistically significant association between our ads measure and the income tax data. This association remains positive and statistically significant when we include both year fixed effects (column 2) and year fixed effects together with borough fixed effects (column 3).

Figure A4: Taxes and Job Ads



(A) Unconstrained (B) Controlling for population

NOTE: Bin scatter graph with each bin summarizing values of several boroughs. The red line represents the linear prediction plot based on the underlying data. Panel A displays the values of (log) borough mentions in newspaper ads and the (log) number of individuals paying income tax under Schedule D in 1870 along two axes. Panel B displays the values of the residuals of (log) borough mentions in newspaper ads and the (log) number of individuals paying income tax under Schedule D in 1870 along two axes when controlling for population (House of Commons (1877)) in the borough along two axes.

Table A6: Individuals paying income tax and Newspaper Job Ads-based variable - Yearly data 1870-1872

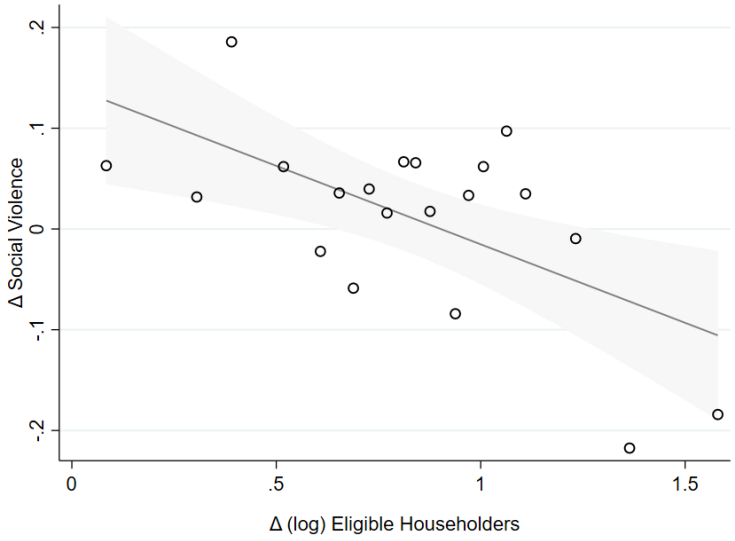
Dep. Variable: (log) Individuals paying income tax$_{it}$	(1)	(2)	(3)
(log) Borough Mentions in Newspaper Ads$_{it}$	0.506***	0.505***	0.0646
	(0.0713)	(0.0714)	(0.0547)
Observations	543	543	543
R-squared	0.284	0.285	0.993
Time FEs	No	Yes	Yes
Borough FEs	No	No	Yes

Notes:The unit of observation is borough i and year t. OLS estimates are reported in all columns. The dependent variable is the (log) number of individuals paying income tax (under Section D) in borough i and year t. The variable *(log) borough Mentions in Newspaper Ads*$_{it}$ corresponds to the number of pages classified as advertisement which contain the word "Wanted" and the name of the borough. Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C. Raw association between our instrument and change in social conflict

Figure A5 below provides a bin scatter illustration of the raw association between Δ (log) eligible householders and social conflict reduction.

Figure A5: Raw association between Δ (log) eligible householders and change in social conflict



NOTE: Bin scatter graph with each bin summarizing values of several boroughs. The red line represents the linear prediction plot based on the underlying data. The x-axis depicts the Δ (log) eligible householders pre and after the Second Reform Act observed in a given borough, while the y-axis plots the difference in (average) social conflict between the period before and after November 1868.

D. Appendix: Robustness Analysis

D.1 Additional Socio-Demographic Controls

In what follows we investigate how sensitive the baseline findings are while controlling for potential confounders. In particular, in Table A7 we explore whether our results hold when we include additional controls (obtained from the Census of 1861 (IPUMS, 2020). We find that controlling for a series of socio-demographic factors does not substantially affect our findings.

Table A7: Democracy and Social Violence - Additional Socio-Demographic Controls

Dep. Variable: Social Violence$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Panel A: OLS Results				
(log) Electorate$_{it}$	-0.139***	-0.139***	-0.142***	-0.117***	-0.135***	-0.133***	-0.137***	-0.150***	-0.130***
	(0.0421)	(0.0446)	(0.0444)	(0.0394)	(0.0418)	(0.0420)	(0.0416)	(0.0428)	(0.0444)
					Panel B: Two-Stage Least Square Results				
(log) Electorate$_{it}$	-0.169***	-0.176***	-0.172***	-0.143**	-0.163***	-0.155***	-0.168***	-0.184***	-0.157**
	(0.0562)	(0.0627)	(0.0579)	(0.0603)	(0.0551)	(0.0568)	(0.0562)	(0.0582)	(0.0682)
1st stage F-Stat	232	194	196	156	228	232	247	243	119
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	No	No	No	No	No	No	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	No	No	No	Yes
*Share Population Working$_i$ * Time FEs*	No	No	No	Yes	No	No	No	No	Yes
*Share Workers in Elementary Occupations$_i$ * Time FEs* 68	No	No	No	No	Yes	No	No	No	Yes
*Sex Ratio$_i$ * Time FEs*	No	No	No	No	No	Yes	No	No	Yes
*(log) Population in 1866$_i$ * Time FEs*	No	No	No	No	No	No	Yes	No	Yes
*Δ Population 1866-1861$_i$ * Time FEs*	No	No	No	No	No	No	No	Yes	Yes

Notes:The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM [2SLS] estimates are reported in Panel A [B]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Controlling for Other Rent-Bands

This Appendix subsection is devoted to a specification where we interact –instead of the Gini Index– the underlying non-parametric rental categories with time dummies. In particular, we interact log householders in the bandwidths [below 4] [10-20) [20-30) [30-40) [40-50) [50-60) [60-70) [70-80) [80-90) [90-100) and [above100) with time dummies (i.e., we interact the rent bands different from our main variable of interest with month fixed effects). The results from this demanding and fine-grained approach to control for the distribution of rents in the borough are presented in Table A8. Our findings prove robust to this demanding sensitivity check.

We also present below Table A9 where the coefficients for all rent-bands (interacted with the post-election period) are displayed. Strikingly, only our rent-band between £4 and £10 is statistically significant.

Table A8: Democracy and Social Violence - Additional Controls for Other Rent-Bands

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) RF	(4) RF	(5) 2SLS	(6) 2SLS
(log) Electors$_{it}$	-0.139***	-0.148**			-0.169***	-0.227**
	(0.0421)	(0.0587)			(0.0562)	(0.0955)
(log) Eligible Householders$_{it}$			-0.146***	-0.151**		
			(0.0483)	(0.0624)		
Observations	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.389	0.428	0.389	0.427	-	-
1st stage F-Stat	-	-	-	-	232	57
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders Below £4$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £10-£20 * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £20-£30$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £30-£40$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £40-£50$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £50-£60$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £60-£70$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £70-£80$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £80-£90$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders between £90-£100$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes
*(log) Householders above £100$_i$ * Time FEs*	No	Yes	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM [2SLS] estimates are reported in columns 1-4 [5-6]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Democracy and Social Conflict - All Rent Bands

Dep. Variable: Social Violence$_{it}$	(1)	(2)	(3)	(4)
(log) Householders below £4$_i$ * *Post-Elections 1868*$_t$	0.00183		0.00153	
	(0.0134)		(0.0130)	
(log) Householders between £4-10$_i$ * *Post-Elections 1868*$_t$	-0.0567*	-0.0567*	-0.0540*	-0.0540*
	(0.0310)	(0.0323)	(0.0286)	(0.0289)
(log) Householders between £10-12$_i$ * *Post-Elections 1868*$_t$	-0.0316			
	(0.0391)			
(log) Householders between £12-14$_i$ * *Post-Elections 1868*$_t$	0.0267			
	(0.0520)			
(log) Householders between £14-16$_i$ * *Post-Elections 1868*$_t$	0.00891			
	(0.0600)			
(log) Householders between £16-18$_i$ * *Post-Elections 1868*$_t$	0.0355			
	(0.0553)			
(log) Householders between £18-20$_i$ * *Post-Elections 1868*$_t$	0.0436			
	(0.0450)			
(log) Householders between £20-30$_i$ * *Post-Elections 1868*$_t$	0.00307			
	(0.0712)			
(log) Householders between £30-40$_i$ * *Post-Elections 1868*$_t$	-0.0913			
	(0.0584)			
(log) Householders between £40-50$_i$ * *Post-Elections 1868*$_t$	-0.00775			
	(0.0698)			
(log) Householders between £50-60$_i$ * *Post-Elections 1868*$_t$	0.00763			
	(0.0429)			
(log) Householders between £60-70$_i$ * *Post-Elections 1868*$_t$	0.0413			
	(0.0435)			
(log) Householders between £70-80$_i$ * *Post-Elections 1868*$_t$	0.00479			
	(0.0340)			
(log) Householders between £80-90$_i$ * *Post-Elections 1868*$_t$	0.0599			
	(0.0394)			
(log) Householders between £90-100$_i$ * *Post-Elections 1868*$_t$	-0.0259			
	(0.0252)			
(log) Householders above £100$_i$ * *Post-Elections 1868*$_t$	-0.0379		-0.0286	
	(0.0230)		(0.0183)	
(log) Householders between £10-50$_i$ * *Post-Elections 1868*$_t$			0.0494	
			(0.0398)	
(log) Householders between £50-100$_i$ * *Post-Elections 1868*$_t$			0.0154	
			(0.0392)	
Observations	4,416	4,416	4,416	4,416
R-squared	0.390	0.413	0.388	0.403
Borough FEs	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Rent Bands$_i$ * *Time FEs*	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM estimates are reported in columns 1-4. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. Each variable corresponds to the log number of householders in the various reported rent bands interacted with a dummy variable equal to 1 in the post election period. In columns (2) and (4) we interact the rent bands different from our main variable of interest with month fixed effect instead of with the dummy variable post. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.3 Controlling for Spatial Fixed Effects

Below we report a robustness check with fixed effects at different levels of aggregation. Table A10 presents the results. Note that the estimates including NUTS3xMonth fixed effects should be interpreted with caution. Indeed, due to the high granularity of the fixed effects we have to drop many singleton observations from the sample. Doing so, we end up with around 3/4 of our initial sample size. For the other cases, we can see that, as we make the fixed effects more granular, the results weaken but often remain statistically significant.

Table A10: Democracy and Social Violence - Adding Spatial Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	OLS	RF	RF	RF	RF	2SLS	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.141***	-0.0972**	-0.109**	-0.0849					-0.175***	-0.117*	-0.108	-0.195*
	(0.0457)	(0.0406)	(0.0469)	(0.0709)					(0.0620)	(0.0601)	(0.0727)	(0.114)
(log) Eligible Householders$_{it}$					-0.158***	-0.101**	-0.0901	-0.168*				
					(0.0549)	(0.0511)	(0.0608)	(0.0948)				
Observations	4,416	4,416	4,296	3,144	4,416	4,416	4,296	3,144	4,416	4,416	4,296	3,144
R-squared	0.397	0.435	0.503	0.564	0.396	0.435	0.502	0.565	-	-	-	-
1st stage F-Stat	-	-	-	-	-	-	-	-	184	108	72	55
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*County$_i$ * Time FEs*	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No
*NUTS 2$_i$ * Time FEs*	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
*NUTS 3$_i$ * Time FEs*	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM [2SLS] estimates are reported in columns 1-8 [9-12]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.4 Controlling for historical electoral changes

Below in Table A11 we investigate if our results are sensitive to controlling for past electoral changes. In particular, we include flexible functional form controls of the electorate changes over the previous elections. Our findings prove robust to the inclusion of these controls.

Table A11: Democracy and Social Violence – Controlling for Past Electoral Changes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	OLS	RF	RF	RF	RF	2SLS	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.141***	-0.131***	-0.125***	-0.125***					-0.175***	-0.169***	-0.174***	-0.174***
	(0.0457)	(0.0450)	(0.0443)	(0.0443)					(0.0620)	(0.0620)	(0.0619)	(0.0619)
(log) Eligible Householders$_{it}$					-0.158***	-0.149***	-0.156***	-0.156***				
					(0.0549)	(0.0539)	(0.0541)	(0.0541)				
Observations	4,416	4,416	4,368	4,368	4,416	4,416	4,368	4,368	4,416	4,416	4,368	4,368
R-squared	0.397	0.401	0.411	0.411	0.396	0.400	0.411	0.411	-	-	-	-
1st stage F-Stat	-	-	-	-	-	-	-	-	184	180	202	202
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Δ Electorate$_{i,1866-1860}$ * Time FEs*	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
*Δ Electorate$_{i,1860-1857}$ * Time FEs*	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
*Δ Electorate$_{i,1857-1832}$ * Time FEs*	No	No	No	No	No	No	No	Yes	No	No	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM [2SLS] estimates are reported in columns 1-8 [9-12]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.5 Using Alternative Inequality Measures

In the current Appendix Section, we construct a borough-level Gini rent inequality index from bined rent categories data using the robust Pareto midpoint estimator proposed by von Hippel, Scarpino and Holas (2016) and Von Hippel, Hunter and Drown (2017). We replicate the baseline regressions when using such alternative inequality measures. While column 1 reproduces the baseline results of columns 3, 6 and 9 of Table 1 for comparison, in columns 2 and 3 of Table A12 we run a variant of our baseline regressions. In particular, we explore whether our results are affected by using *Mehran* and *Piesch indexes* of inequality. Similarly, columns 4 and 5 show the results obtained using the *Kakwani* and *Theil indexes*. Finally, in the last two columns we use as measures of rent inequality a *Generalized Entropy Index* and the *Mean-Log Deviation* of rents observed in the borough. It turns out that our results are virtually unchanged when controlling for these alternative inequality measures.

Table A12: Democracy and Social Violence - Alternative Inequality Indexes

Dep. Variable: Social Violence$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: OLS Results			
(Log) Electorate$_{it}$	-0.141***	-0.137***	-0.142***	-0.141***	-0.141***	-0.139***	-0.140***
	(0.0457)	(0.0438)	(0.0469)	(0.0469)	(0.0491)	(0.0421)	(0.0462)
				Panel B: Reduced Form Results			
(Log) Eligible Householders$_{it}$	-0.158***	-0.155***	-0.158***	-0.158***	-0.155***	-0.147***	-0.158***
	(0.0549)	(0.0542)	(0.0554)	(0.0554)	(0.0565)	(0.0482)	(0.0551)
				Panel C: Two-Stage Least Square Results			
(Log) Electorate$_{it}$	-0.175***	-0.170***	-0.178***	-0.177***	-0.181***	-0.173***	-0.176***
	(0.0620)	(0.0605)	(0.0637)	(0.0636)	(0.0676)	(0.0607)	(0.0624)
1st stage F-Stat	184	199	171	172	148	195	181
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	Yes	No	No	No	No	No	No
*Mehran Index Rents$_i$ * Time FEs*	No	Yes	No	No	No	No	No
*Piesch Index Rents$_i$ * Time FEs*	No	No	Yes	No	No	No	No
*Kakwani Index Rents$_i$ * Time FEs*	No	No	No	Yes	No	No	No
*Theil Index Rents$_i$ * Time FEs*	No	No	No	No	Yes	No	No
*Generalized Entropy Measure Rents$_i$ * Time FEs*	No	No	No	No	No	Yes	No
Mean Log Deviation Rents$_i$ Time FEs*	No	No	No	No	No	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM [2SLS] estimates are reported in Panels A and B [Panel C]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*) Additional controls have been constructed using the 1861 Population Census conducted by the Secretary of State of the United Kingdom (IPUMS (2020)). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.6 Alternative Instruments

Like in the previous Appendix Section, here we assess our instrument sensitivity. In particular, Table A13 displays the results.

Columns 1 and 2 reproduce the baseline OLS results of Table 1, in the goal of providing a quantitative benchmark. Similarly, columns 3 and 4 reproduce our baseline 2SLS results from Table 1. In contrast, the results in columns 5-6 are obtained when using an alternative instrument where the value of the variable *Eligible householders* in the period after the election it is equal to the (log) number of householders above £10 and the log of the number of householders living in a house with rental value between 0 and £10.

The results for this alternative instrument (which assumes that after the reform all householders can vote) are very similar to those for the baseline instrument (which accounts for the evidence that after the reform typically only householders occupying property of rental value above £4 can vote). This is reassuring, as it suggests that our results are not sensitive to the threshold of £4.

Table A13: Democracy and Social Violence - Alternative Instruments

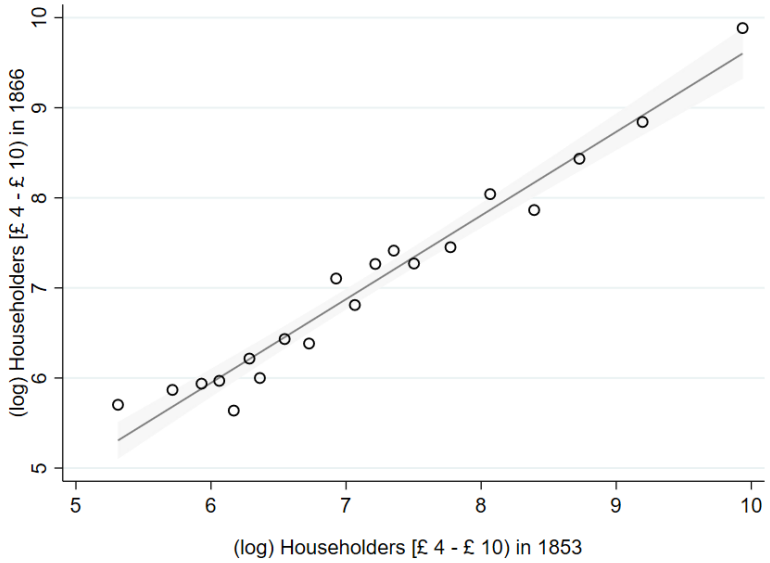
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	2SLS	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.139***	-0.141***	-0.169***	-0.175***	-0.146***	-0.155***
	(0.0421)	(0.0457)	(0.0562)	(0.0620)	(0.0493)	(0.0566)
Observations	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.389	0.397	-	-	-	-
1st stage F-Stat	-	-	232	184	237	209
Instrumental Variable	-		4 £ - Over in 1866		All rents in 1866	
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes
Rent Inequality$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-2 [3-6]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i in columns 1-2. The variable *(log) Electorate*$_{it}$ is instrumented in columns 3-4 with the standard variable *(log) Eligible Householders*$_{it}$ corresponding to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. In columns 5-6 the *(log) Electorate*$_{it}$ is instrumented with the variable *(log) Eligible Householders* which in the period after the election it is equal to the (log) number of householders above £10 and the log of the number of householders living in a house with rental value between 0 and £10. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.7 Alternative (i.e. older) Rents Data

Furthermore, there could also be concerns about strategic rent manipulation around the threshold of £10. First of all, it is important to note that, reassuringly, in the descriptive statistics of the rent distribution (see Figure A1 above) we did not detect any indication of a jump between the rent categories of "between 8 and 10" and "between 10 and 12", respectively, hence there does not appear to be a plausible risk of strategic rent manipulation.
Nevertheless, we perform a battery of robustness exercises where we use information related to the distribution of rents more than ten years before the reform act (i.e. in 1853). By exploiting an older distribution, we are able to mitigate concerns related to a potential sorting of valuation of properties around the election period. It is worth noting that the number of householders between £4 and £10 has remained fairly stable over time. Looking that the numbers observed in 1853 (House of Commons, 1860*a*) correspond with the numbers in 1866, it appears that there have been limited changes and these changes do not go in a specific direction (see Figure A6 below).
In Table A14 below we explicitly control for any pre-reform changes in the rent distribution. In particular, we replicate our baseline estimates but add as a further control the growth rate of householders in the bandwidth between £4 and £10 between 1866 and 1853 interacted with time dummies. The first three columns of this Table contain our reduced form estimates while the last three columns display 2SLS results. The inclusion of this additional set of control variables leaves our estimates substantially unchanged.
Finally, in Table A15 we use the rents distribution of householders in 1853 (instead of 1868) for the intention to treat (ITT) analysis and to construct our instrument. This serves the purpose of addressing concerns about strategic rent manipulation. Also in the case of this robustness check, our results remain stable and provide support for the pacifying effects of franchise extension.

Figure A6: Householders with rents between [£4-£10] in 1866 and in 1853



NOTE: The figure displays the values of (log) number of householders with rents between [£4-£10] observed in 1866 and the (log) number of householders with rents between [£4-£10] observed in 1853 along two axes.

Table A14: Democracy and Social Violence - Controlling for $\Delta_{1866-1853}$ Householders [£4 - £10]

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	2SLS	2SLS	2SLS
(log) Eligible Householders$_{it}$	-0.146***	-0.141***	-0.149***			
	(0.0483)	(0.0533)	(0.0559)			
(log) Electorate$_{it}$				-0.169***	-0.160***	-0.159***
				(0.0562)	(0.0611)	(0.0602)
Observations	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.389	0.392	0.399	-	-	-
1st stage F-Stat	-	-	-	232	175	180
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	No	Yes	No	No	Yes
Rent Inequality$_i$ ** Time FEs*	No	No	Yes	No	No	Yes
$\Delta_{1866-1853}$ *Householders [£4 - £10] * Time FEs*	No	Yes	Yes	No	Yes	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-3 [4-6]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*) and House of Commons (1860*b*). Columns 3 and 6 include flexible controls for the growth rate of householders with rents between [$£$4-$£$10) observed between 1866 and 1853. Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Democracy and Social Violence - Using Rental Data for 1853

Dep. Variable: Social Violence$_{it}$	(1) RF	(2) RF	(3) 2SLS	(4) 2SLS	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS
(log) Electorate$_{it}$			-0.171***	-0.169**			-0.173**	-0.164*
			(0.0624)	(0.0750)			(0.0781)	(0.0935)
(log) Eligible Householders [Predicted]$_{it}$	-0.133***	-0.132**						
	(0.0479)	(0.0604)						
(log) Eligible Householders [1853]$_{it}$					-0.120**	-0.128*		
					(0.0541)	(0.0749)		
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.387	0.394	-	-	0.387	0.394	-	-
1st stage F-Stat	-	-	88	55	-	-	75	50
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	No	No	No	Yes	No	No
Rent Inequality $_i$ ** Time FEs*	No	Yes	No	No	No	Yes	No	No
Sample Mean	.237	.237	.237	.237	.237	.237	.237	.237

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-2-5-6 [3-4-7-8]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders [Predicted]*$_{it}$ is obtained using the number of householders with rental value above $£10$ obtained from the rental statistics for the year 1866 and the number of householders living in a house with rental value between $£4$ and 10 (who were previously banned from voting but enfranchised in the Second Reform Act) obtained from the rental statistics for the year 1853. The variable *(log) Eligible Householders [1853]*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868 using the rental-statistics observed for the year 1853. In columns 3 and 4 (7 and 8) are reported 2SLS results obtained with *(log) Eligible Householders [Predicted]*$_{it}$ (*(log) Eligible Householders [1853]*$_{it}$) as instrument. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*) and House of Commons (1860*b*). Columns 3 and 6 include flexible controls for the growth rate of householders with rents between [$£4$-$£10$) observed between 1866 and 1853. Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.8 Alternative Dependent Variables

Below we explore the relationship between electors and social violence looking at the number of occurrences in a month. In particular, we use as a dependent variable the (log) number of events in a month (Table A16) and the inverse hyperbolic sine transformation (Table A17), respectively. Further, we run a Poisson regression in Table A18. In all cases our results remain statistically significant over all specifications.

Table A16: Democracy and Social Violence - Alternative Dependent Variable - Intensive Margin [1]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.194***	-0.200***	-0.203***				-0.231**	-0.255**	-0.252**
	(0.0654)	(0.0676)	(0.0702)				(0.0914)	(0.100)	(0.0979)
(log) Eligible Householders$_{it}$				-0.200**	-0.228**	-0.227**			
				(0.0791)	(0.0887)	(0.0875)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.536	0.538	0.543	0.535	0.537	0.542	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.276	.276	.276	.276	.276	.276	.276	.276	.276

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS [2SLS] estimates are reported in columns 1-6 [7-9]. The dependent variable is the (log+1) number of violent events observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Democracy and Social Violence - Alternative Dependent Variable - Intensive Margin [2]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.245***	-0.253***	-0.257***				-0.291**	-0.321**	-0.318***
	(0.0821)	(0.0850)	(0.0881)				(0.114)	(0.125)	(0.122)
(log) Eligible Householders$_{it}$				-0.252**	-0.288**	-0.287***			
				(0.0988)	(0.111)	(0.109)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.534	0.536	0.541	0.533	0.536	0.540	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.353	.353	.353	.353	.353	.353	.353	.353	.353

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is the inverse hyperbolic sine transformation of violent events observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A18: Democracy and Social Violence - Alternative Dependent Variable - Intensive Margin [3]

Dep. Variable: Social Violence Events$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)
(log) Electorate$_{it}$	-0.598**	-0.813***	-0.814***			
	(0.246)	(0.261)	(0.232)			
(log) Eligible Householders				-0.542*	-0.894***	-0.928***
				(0.280)	(0.331)	(0.300)
Observations	3,816	3,816	3,816	3,816	3,816	3,816
1st stage F-Stat	-	-	-	-	-	-
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Gross Rental * Month FEs	No	Yes	Yes	No	Yes	Yes
Gini Gross Rental * Month FEs	No	No	Yes	No	No	Yes
Sample Mean	.833	.833	.833	.833	.833	.833

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. Poisson estimates are reported in columns 1-6. The dependent variable is the number of violent events observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.9 Logit

As discussed in Section 5.1, in Table A19 below we replicate our main baseline specifications using conditional logit regressions instead of the linear probability model that we have used throughout the paper. Like for the baseline analysis, we continue to find a statistically significant social conflict-reducing effect of franchise extension.

Table A19: Democracy and Social Violence - Fixed Effect Logit Estimator

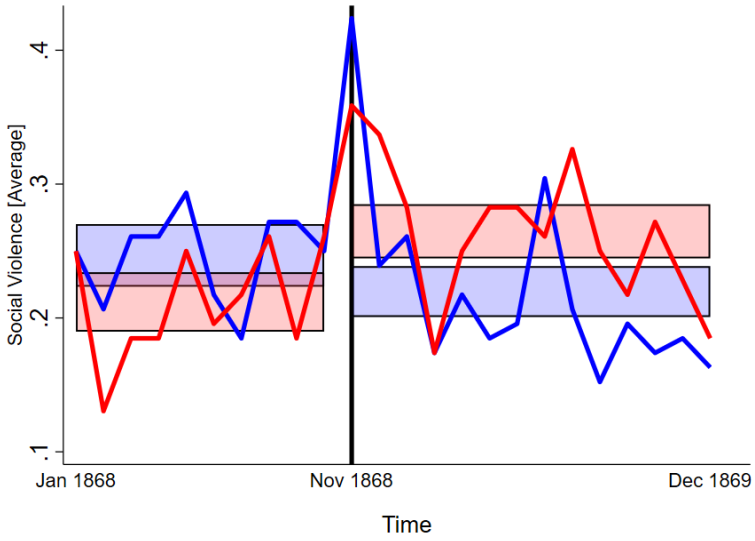
Dep. Variable: Social Violence$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)
(log) Electorate$_{it}$	-1.114***	-1.084***	-1.113***			
	(0.310)	(0.338)	(0.347)			
(log) Eligible Householders$_{it}$				-1.103***	-1.215***	-1.208***
				(0.358)	(0.457)	(0.459)
Observations	3,768	3,768	3,768	3,768	3,768	3,768
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ ** Time FEs*	No	No	Yes	No	No	Yes
Sample Mean	.265	.265	.265	.265	.265	.265

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. Fixed effects logit estimates are reported in all columns. The dependent variable is the number of violent events observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.10 Intensity of Enfranchisement and Evolution of Social Violence

As a first way of tackling the issue of a common pre-trend, in the current Appendix Section we study graphically the evolution of both low versus high enfranchisement boroughs. In particular, in the spirit of an event study, in Figure A7 below we compare the evolution of social violence for below-median enfranchisement boroughs (in red) versus boroughs with above-median franchise extension (blue line). We find that before the election the level of social violence in the two types of boroughs moved largely in parallel, whereas post-reform the more enfranchised boroughs displayed on average a lower level of social violence.

Figure A7: Intensity of Δ Electorate and Social Violence



NOTE: The figure displays the evolution of social violence in low intensity (red line) and high intensity (blue line) municipalities. Red (blue) area shows the average social conflict (and the 90% confidence intervals) observed over the two periods. The sample covers 184 boroughs over the period 1868-1869. Low (high) boroughs are defined as all municipalities where the increase in electorate is below the median (above the median). The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1.

D.11 Synthetic Control Method

As a next step for addressing worries about common pre-trends in high versus low enfranchisement boroughs, the current Appendix Section presents results using the Synthetic Control Method (SCM). This constitutes a transparent method of choosing counterfactual units, and has recently been applied e.g. in Abadie and Gardeazabal (2003), Billmeier and Nannicini (2013).

In our setting, we define treated and potential counterfactual units based on the intensity of the treatment. That is, the potential control units for a given borough are all boroughs where the enfranchisement brought by the Second Reform Act was lower than for the unit of interest.[31]

For each borough, we apply the synthetic algorithm to construct a counterfactual unit as a weighted combination of a group of potential counterfactual units. Weights are selected in order to approximate the incidence of social conflict events of the unit in question prior to reform, using a transparent data-driven procedure. To ensure that the results are not driven by the inclusion of any particular district and to assess statistical significance of our estimates, we replicate this procedure using 500 different groups of potential counterfactuals, where each counterfactual group is computed randomly by drawing on two-thirds of all control districts.[32]

In order to assess the total effect of the reform at the aggregate level we combine all *treated* boroughs and the corresponding synthetic counterfactual observations. In doing so, we are able to compare the actual incidence of social violence observed in the United Kingdom with the distribution of violence observed in the 500 aggregate synthetic counterfactual units.

Panel A in Figure A8 displays the results obtained using the Synthetic Control Method (SCM). We can see that the treated group (with high enfranchisement, represented by the slim solid line) follows –by construction– an extremely similar path before the electoral reform, but displays systematically lower levels of social conflict post-reform. Similar findings are obtained when we use our instrumental variable as a*treatment variable of interest* (Δ *Eligible Householders*$_i$) [Panel B of Figure A8].[33]

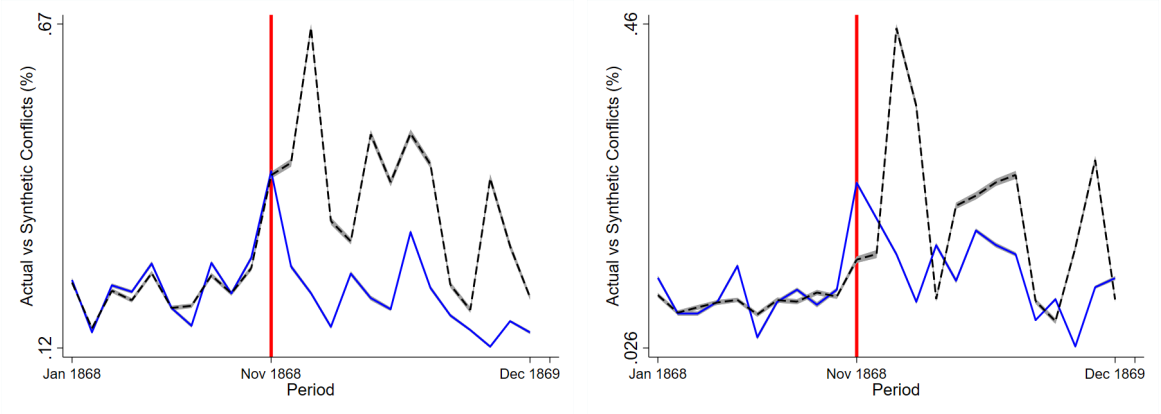
To rule out that this result is obtained "by chance" or due to some "mechanical" measurement

---

[31] In this Section, our explanatory variable of interest is the variable Δ *Electorate*$_i$ is computed as $log((Electors\ Post\ -\ Reform)/(Electors\ Pre\ -\ Reform))_i$ where $(Electors\ Post\ -\ Reform)_i$ and $(Electors\ Pre\ -\ Reform)_i$ correspond to the electorate pre and after the Second Reform Act in a borough i, respectively. We implement the synthetic control method for all districts where the delta in the number of electors is above the $25th$ percentile. This is due to the fact that the synthetic algorithm requires a certain number of potential counterfactual units. Allowing for meaningful difference between treated and controls units, we select as potential counterfactuals only those with a value of Δ Electorate less than half of the one observed in the corresponding *treated* unit.

[32] Further details on the use of subsampling methods as inferential tools for synthetic control estimators are pre-

Figure A8: Democracy and Social Violence - Synthetic Control Method



(A) Treat. Var.: Δ *Electorate*$_i$

(B) Treat. Var.: Δ *Eligible Householders*$_i$

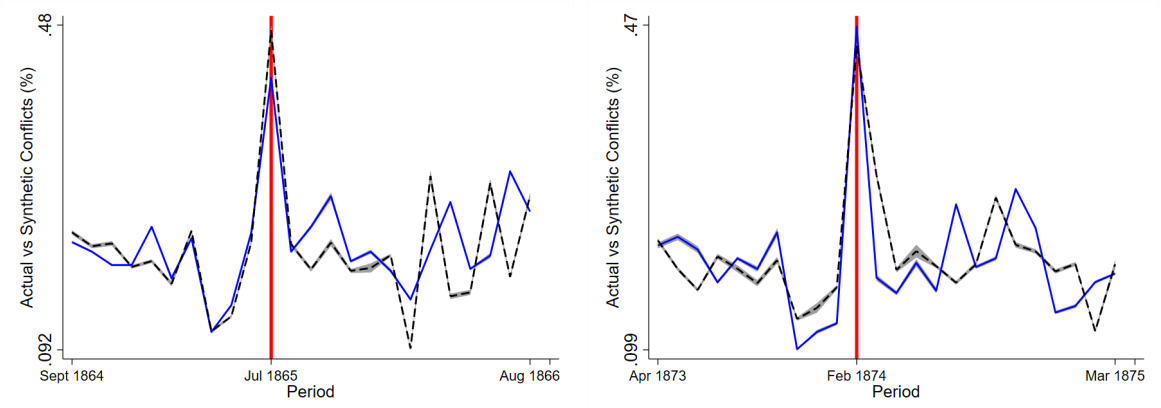
NOTE - Left Panel: The solid line corresponds to the actual average incidence of social violence observed in all boroughs, while the dashed line captures the average incidence of violence obtained from synthetic counterfactuals. The dark grey area around the dashed line indicates the 99% confidence interval. Each synthetic unit was computed as a weighted average of randomly drawn group districts where the intensity of the enfranchisement due to the passage of the Reform was lower than in the district of interest. Weights are selected according to the incidence of social conflict events of the unit in question prior to the elections of 1868. Right Panel: The solid line corresponds to the actual average incidence of social violence observed in all boroughs, while the dashed line captures the average incidence of violence obtained from synthetic counterfactuals. The dark grey area around the dashed line indicates the 99% confidence interval. Each synthetic unit was computed as a weighted average of randomly drawn group districts where the value of the variable Δ *Eligible Householders*$_i$ was lower than in the borough of interest. Weights are selected according to the incidence of social conflict events of the unit in question prior to the elections of 1868.

error, we perform a placebo SCM analysis on other election years where no large-scale franchise extension occurred and where accordingly we do not expect any effects. The results are displayed in Figure A9, where, as expected, we do not perceive any systematic differences between the treatment and control group after these "placebo" elections.

sented in Saia (2017).

[33]The variable Δ *Householders Below Threshold*$_i$ is computed as $log((Number\ Householders\ £4 - 10 + Number\ Householders\ Above\ £10)/(Number\ Householders\ Above\ £10))_i$, where $(Number\ Householders\ £4 - 10)_i$ and $(Number\ Householders\ Above\ £10)_i$ correspond to the number of householders living in a house with rental value between £4 and 10 (who were previously were banned from voting but enfranchised after the Second Reform Act) and the number of householders with rental value above £10 (who could already vote before the reform) in a borough i, respectively.

Figure A9: Democracy and Social Violence - Synthetic Control Method - Placebo Elections



(A) Elections 1865 (B) Elections 1874

NOTE: Left [Right] Panel: The solid line corresponds to the actual average incidence of social violence observed in all boroughs, while the dashed line captures the average incidence of violence obtained from synthetic counterfactuals. The dark grey area around the dashed line indicates the 99% confidence interval. Each synthetic unit was computed as a weighted average of randomly drawn group districts where the intensity of the enfranchisement due to the passage of the Reform was lower than in the district of interest. Weights are selected according to the incidence of social conflict events of the unit in question prior to the elections of 1865 [1874].

D.12 Alternative levels of clustering

As mentioned in Section 5.1 of the main text, Tables A20 to A22 below show that the conclusions of the statistical inference continue to hold when we allow for standard errors to be clustered at alternative levels. In particular, Table A20 displays the results with several different clustering levels, among others for robust standard errors and for standard errors that are two-way clustered at the borough and month level. Table A21 allows for spatial correlation in error terms using Conley (1999)'s approach and a range of different reference distances for spatial decay. Finally, in Table A22 we cluster standard errors at the present day NUTS3 level. Throughout the various clustering levels, the statistical inference is similar, and we continue to find a statistically significant effect of enfranchisement on reducing social conflict.

Table A20: Democracy and Social Violence - Alternative Levels of Clustering

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.139***	-0.139***	-0.141***				-0.169***	-0.176***	-0.175***
Cluster City	(0.0421)	(0.0446)	(0.0457)				(0.0562)	(0.0627)	(0.0620)
Cluster Robust	(0.0274)	(0.0289)	(0.0291)				(0.0371)	(0.0395)	(0.0394)
Double Cluster Borough-Month	(0.0439)	(0.0459)	(0.0467)				(0.0614)	(0.0660)	(0.0653)
(log) Eligible Householders$_{it}$				-0.146***	-0.158***	-0.158***			
Cluster Borough				(0.0483)	(0.0550)	(0.0549)			
Cluster Robust				(0.0321)	(0.0353)	(0.0354)			
Double Cluster Borough-Month				(0.0528)	(0.0580)	(0.0579)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.389	0.392	0.397	0.389	0.392	0.396	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Standard error computed using clustered at clustered at alternative levels are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A21: Democracy and Social Violence - Alternative Levels of Spatial Clustering

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.139***	-0.139***	-0.141***				-0.169***	-0.176***	-0.175***
Cluster Borough	(0.0421)	(0.0446)	(0.0457)				(0.0562)	(0.0627)	(0.0620)
Spatial Cluster 25km	(0.0291)	(0.0306)	(0.0307)				(0.0383)	(0.0416)	(0.0409)
Spatial Cluster 50km	(0.0310)	(0.0326)	(0.0325)				(0.0402)	(0.0432)	(0.0422)
Spatial Cluster 100km	(0.0308)	(0.0324)	(0.0321)				(0.0391)	(0.0413)	(0.0400)
Spatial Cluster 150km	(0.0288)	(0.0304)	(0.0295)				(0.0398)	(0.0422)	(0.0403)
(log) Eligible Householders$_{it}$				-0.146***	-0.158***	-0.158***			
Cluster Borough				(0.0483)	(0.0550)	(0.0549)			
Spatial Cluster 25km				(0.0336)	(0.0376)	(0.0373)			
Spatial Cluster 50km				(0.0357)	(0.0397)	(0.0392)			
Spatial Cluster 100km				(0.0346)	(0.0376)	(0.0368)			
Spatial Cluster 150km				(0.0347)	(0.0378)	(0.0365)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.389	0.392	0.397	0.389	0.392	0.396	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Standard error computed using clustered at clustered at alternative (spatial) levels are reported in parenthesis. Spatial standard errors are computed using the *acreg* command coded by Colella, Lalive, Sakalli and Thoenig (2019). Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A22: Democracy and Social Conflict - Clustering at NUTS3 Level

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) RF	(4) RF	(5) 2SLS	(6) 2SLS
(log) Electorate$_{it}$	-0.139***	-0.141**			-0.169***	-0.175**
	(0.0508)	(0.0550)			(0.0640)	(0.0705)
(log) Eligible Householders$_{it}$			-0.146**	-0.158**		
			(0.0576)	(0.0653)		
Observations	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.389	0.397	0.389	0.396	-	-
1st stage F-Stat	-	-	-	-	193	148
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	No	Yes	No	Yes
Rent Inequality$_i$ * *Time FEs*	No	Yes	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-4 [5-6]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Standard errors are clustered at administrative NUTS3 level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.13 Outliers and sample composition

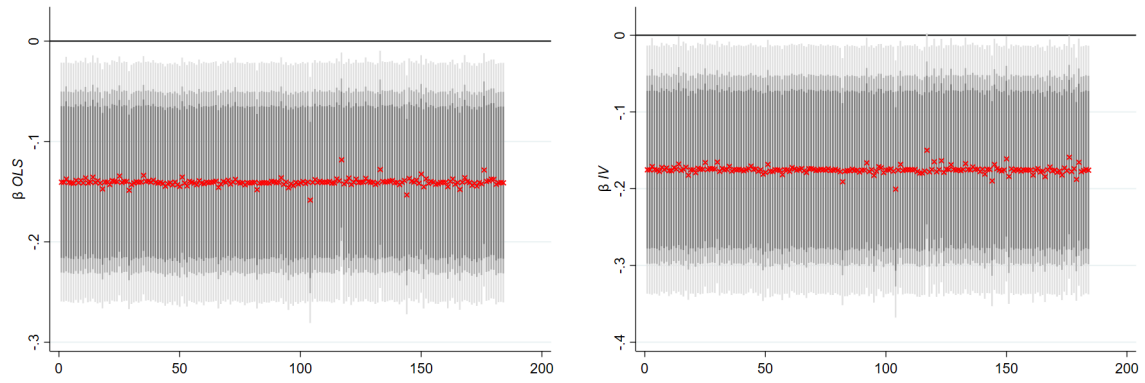
As discussed earlier in Section 5.1, we display below a series of robustness results when removing data from one borough at a time, when discarding social conflict-events from one newspaper at a time or when randomly eliminating days in our time-window.

Firstly, we replicate our baseline regressions when dropping one borough at the time. Panels A and B in Figure A10 display the distribution of coefficients estimated using baseline regressions presented in columns 3 and 9 of Table 1, respectively. The corresponding point estimates are hardly affected when removing units from our sample and always reach conventional levels of statistical significance.

Secondly, we explore whether our results hold if we drop one newspaper source at the time. Corresponding results are reported in Figure A11. It turns out that the results remain very similar and we still find a statistically significant social conflict-reducing impact of franchise extension.

Finally, we assess whether our results are robust when only a subsample of days in our sample period are used. To this end we carry out a Monte Carlo analysis with 1,000 repetitions where for each draw only two-thirds of days are kept in our sample period. Panels A and B in Figure A12 display the distribution of coefficients estimated using the baseline specifications of columns 3 and 9 of Table 1, respectively. In both cases, point estimates of the coefficient of interest appear to be fairly stable to the sample removal exercises and are consistent with the baseline estimates reported in the main text.

Figure A10: Democracy and Social Violence - Dropping One borough at the Time

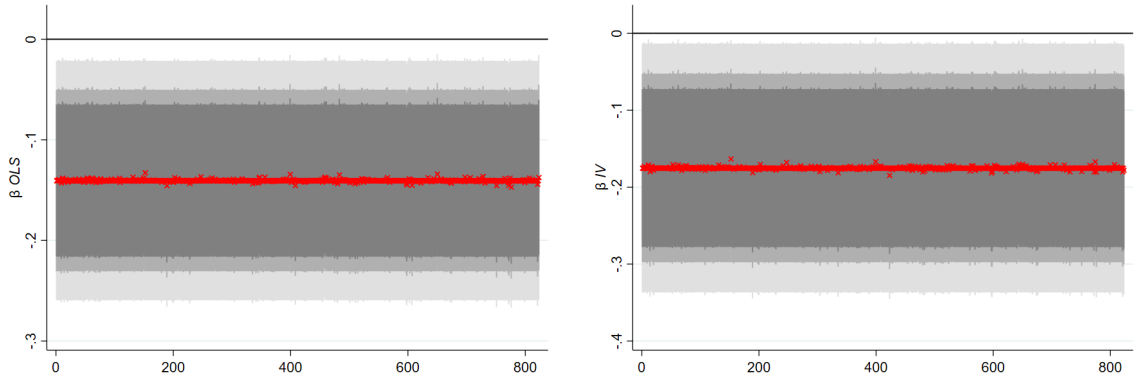


(A) LPM Estimates

(B) 2SLS Estimates

NOTE: Panel A [Panel B] displays the distribution of coefficients estimated for the variable *(log) Electorate*$_{it}$ using the specification of column 3 [column 9] of Table 1 obtained by removing one borough at the time from the sample. Red dots indicate the point estimate of the coefficient of interest. Grey bars depict confidence intervals [the lighter the bar, the higher is the confidence threshold (i.e., light grey indicates 99% C.I.)].

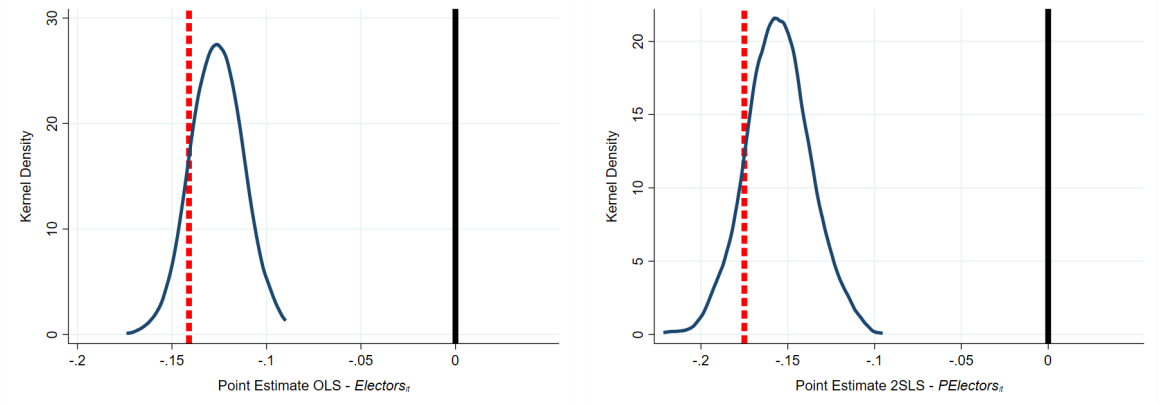
Figure A11: Democracy and Social Violence - Dropping One Newspaper at the Time



(A) LPM Estimates (B) 2SLS Estimates

NOTE: Panel A [Panel B] displays the distribution of coefficients estimated for the variable *(log) Electorate*$_{it}$ using the specification of column 3 [column 9] of Table 1 obtained by removing one newspaper at the time from the sample. Red dots indicate the point estimate of the coefficient of interest. Grey bars depict confidence intervals [the lighter the bar, the higher is the confidence threshold (i.e., light grey indicates 99% C.I.)].

Figure A12: Distribution of coefficients estimated using 1,000 sample periods



(A) LPM Estimates (B) 2SLS Estimates

NOTE: Panel A [Panel B] displays the distribution of 1,000 coefficients estimated for the variable *(log) Electorate*$_{it}$ using the specification of column 3 [column 9] of Table 1 obtained by (randomly) removing one-third of days in our sample period. Red line indicates the point estimate of the coefficient of interest obtained with the full sample.

D.14 Alternative Time Frequencies

In order to investigate whether our results hinge on the use of monthly time units, we construct the dataset in two alternative ways, i) at the weekly level, and ii) with simply one pre- and one post-reform period. The results reported in A23 are the ones obtained using a weekly-level panel, while Table A24 depicts the findings for a two period (pre-post) panel. In both cases we continue to find a strong and significant social conflict-reducing impact of franchise extension.

Table A23: Democracy and Social Violence - Weekly-Level Panel

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.0568**	-0.0606***	-0.0615***				-0.0618**	-0.0716**	-0.0710**
	(0.0221)	(0.0228)	(0.0235)				(0.0301)	(0.0325)	(0.0318)
(log) Eligible Householders$_{it}$				-0.0536**	-0.0642**	-0.0639**			
				(0.0262)	(0.0290)	(0.0286)			
Observations	19,136	19,136	19,136	19,136	19,136	19,136	19,136	19,136	19,136
R-squared	0.291	0.295	0.299	0.290	0.294	0.298	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality $_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.098	.098	.098	.098	.098	.098	.098	.098	.098

Notes: The unit of observation is borough i and week w. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and week t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: Democracy and Social Violence - Pre-Post Panel

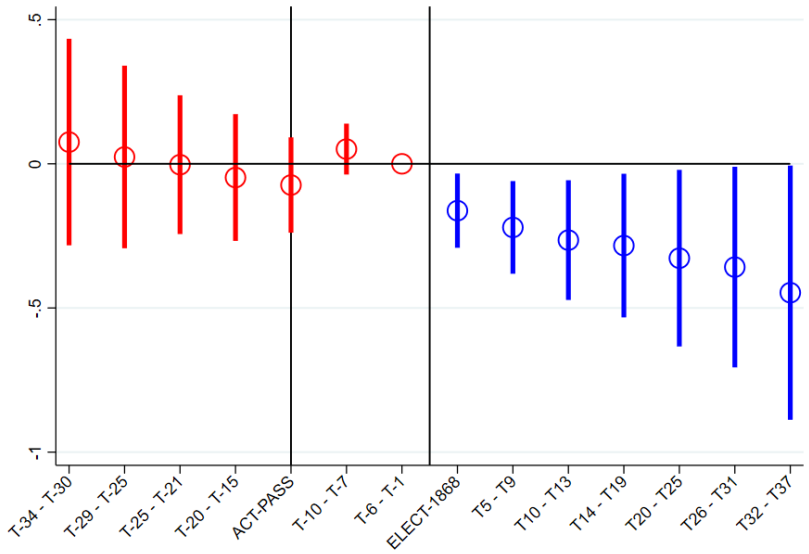
Dep. Variable: Social Violence$_{it}$	(1) Avg. Months OLS	(2) Log Months OLS	(3) Log Events OLS	(4) Avg. Months RF	(5) Log Months RF	(6) Log Events RF	(7) Avg. Months 2SLS	(8) Log Months 2SLS	(9) Log Events 2SLS
(log) Electorate$_{it}$	-0.141***	-0.313	-0.229*				-0.175***	-0.496*	-0.352*
	(0.0456)	(0.197)	(0.137)				(0.0619)	(0.264)	(0.192)
(log) Eligible Householders$_{it}$				-0.158***	-0.447*	-0.317*			
				(0.0547)	(0.233)	(0.169)			
Observations	368	368	368	368	368	368	368	368	368
R-squared	0.899	0.858	0.840	0.897	0.858	0.841	-	-	-
1st stage F-Stat	-	-	-	-	-	-	183	183	183
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rent Inequality$_i$ ** Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Sample Mean	.236	1.303	.993	.236	1.303	.993	.236	1.303	.993

Notes: The unit of observation is borough i and period t, where t (t-1) represents the period before (after) the elections of 1868. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. In columns 1-4-7 [2-5-8] (3-6-9) dependent variable is the average number of months with social violence in borough i and period t [the (log+1) number of months with violent events in borough i and period t] (the (log+1) number of violent events in borough i and period t). The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.15 Alternative specifications and alternative sample coverage

In this Appendix, we assess whether results i) hold when we allow for a borough-specific trend and ii) are robust to alternative sample coverage. Results presented in Table A25 show that the pacifying effect of enfranchisement is also found when we include a borough-specific trend (to remove any borough-specific trend in violence) and we expand the coverage to include both longer periods before and after our baseline sample. We also study the evolution of the expansion of the electorate over time, but this time by slicing the sample in several subperiods, using the specification (and the sample period) adopted in Column 6 of Table A25. The corresponding results are presented in Figure A13. As expected, before the reform future enfranchisement does not matter, resulting in small coefficients (far from being statistically significant and with lower magnitude, which is consistent with the common pre-trend assumption). Furthermore, we find that the *de jure* passing of the law was not enough to reduce social conflict, and we also do not observe any change in social conflict events between the passing of the act and the first election thereafter in 1868. In contrast, after the 1868 election the level of social conflict drops.

Figure A13: Democracy and Social Conflict - Leads and Lags - Extended Sample



NOTE: The figure displays the coefficients of estimates of leads and lags of the variable *(log) Electorate Post*$_{it}$ (i.e. the number of electors post reform). In the specification we also control for *(log) Electorate Pre*$_{it}$ (i.e. the number of electors pre reform). These variables are interacted with time dummies taking value 1 for each time-window. Time-windows are displayed on the horizontal axis (omitted period is [T-6:T-1]). Estimates are obtained including interactions of average rents and rent inequality with monthly time dummies. Moreover we control for borough specific time trends. In this way we obtain leads and lags specifications comparable with Column 6 *Panel A* of Table A25. The dependent variable is *Social Violence*$_{it}$.

Table A25: Democracy and Social Conflict - Extended Sample

Dep. Variable: Social Violence$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS Results			
(Log) Electorate$_{it}$	-0.141***	-0.109*	-0.200***	-0.159***	-0.151***	-0.115**
	(0.0457)	(0.0637)	(0.0630)	(0.0545)	(0.0519)	(0.0481)
			Panel B: Two-Stage Least Square Results			
(Log) Electorate$_{it}$	-0.175***	-0.0475	-0.218***	-0.185**	-0.174**	-0.117*
	(0.0620)	(0.0824)	(0.0805)	(0.0731)	(0.0716)	(0.0644)
1st stage F-Stat	184	176	178	180	181	181
Observations	4,416	4,416	6,624	8,832	11,040	13,248
Sample Period	*Baseline*	*Baseline*	*1867-1869*	*1866-1869*	*1866-1870*	*1866-1871*
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Borough-Specific Linear Trend	No	Yes	Yes	Yes	Yes	Yes

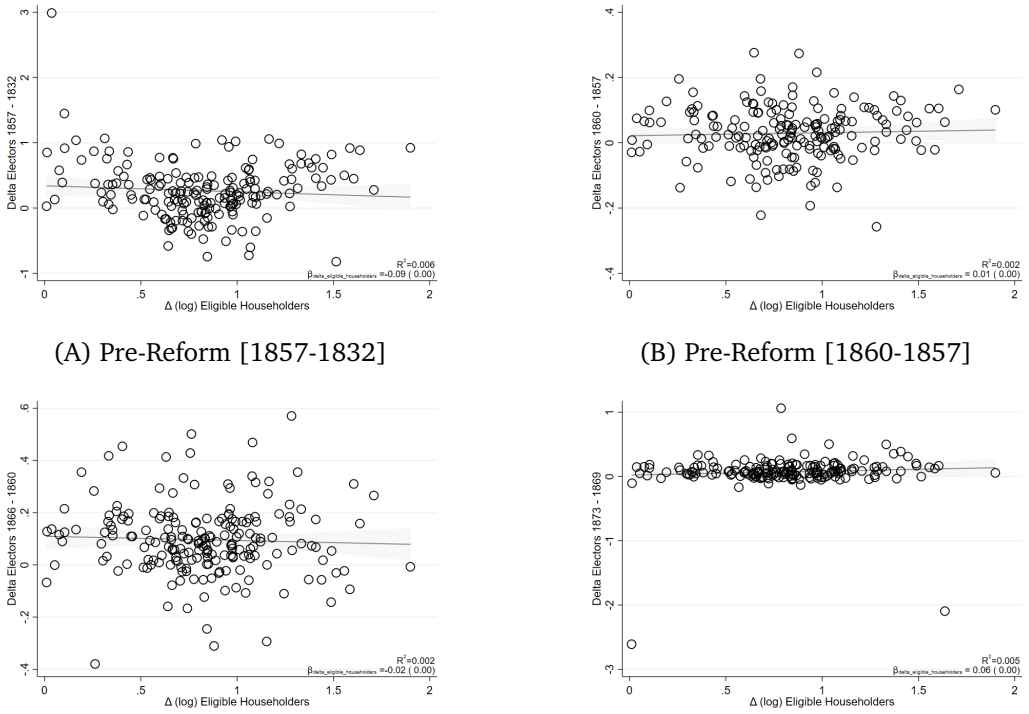
Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over different time periods (reported in the table, the *Baseline* period corresponds to the time period used in Table 1). LPM estimates are reported in *Panel A* while 2SLS estimates are in *Panel B*. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. In *Panel B* the variable *(log) Electorate*$_{it}$ is instrumented using the variable *(log) Eligible Householders*$_{it}$ which corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). In column 1 we report the baseline results corresponding to columns 3 and 9 of table 1. In column 2 we add town specific time trend. In column 3 we extend the sample backward adding all the months of year 1867. In column 4 we add year 1866, while in columns 5 and 6 we extend the sample also forward adding years 1870 and 1871 respectively. In all specifications we include all the standard controls present in column 3 of table 1. Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.16 Placebo IVs

One potential worry could be that our instrument "mechanically" picks up some measurement error or borough characteristics unrelated to electoral reform. To investigate this possibility, we start by studying in Figure A14 whether our IV also predicts electorate changes outside large-scale reforms (which it should not, if our identification strategy is valid). Reassuringly, the change in eligible renters within the 4 to 10 pounds bracket does not correlate with the change in electors neither in the pre-reform period nor in the period after the reform – as expected, it only explains the change in the electorate in 1868 when the Second Reform Act was implemented.

As a next step, we replicate the main analysis using the period around the election in 1865 [elections prior to the introduction of the Second Reform Act] and in 1874 [after the reform]. In other words, we investigate whether our instrumental variable had an effect during the elections of 65 and 74 (which, again, it should not, if our identification strategy is valid). In Figure A15 we display the reduced-form coefficients obtained using the elections of 1865 [11–24 July 1865], 1868 and 1874 [31 January – 17 February 1874]. We take time-windows across different elections and compare the effect of our instrument on social violence in the post-election periods for these various elections. Reassuringly, the displayed coefficients are statistically significant only in the election of 1868.

Figure A14: Δ Electors and Δ Eligible Householders



(A) Pre-Reform [1857-1832]

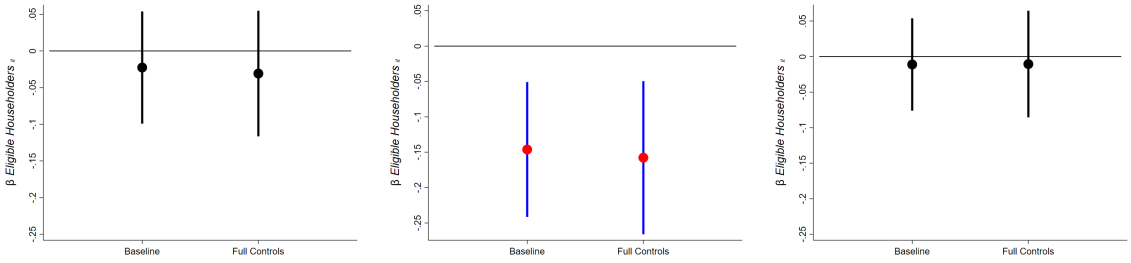
(B) Pre-Reform [1860-1857]

(C) Pre-Reform [1866-1860]

(D) Post-Reform [1873-1869]

NOTE: Panel A displays the values of Δ Electorate between 1832 and 1857 and Δ eligible householders in 1866 in English boroughs along two axes. Panel B displays the values of Δ Electorate between 1857 and 1860 and Δ eligible householders in 1866 in English boroughs along two axes. Panel C displays the values of Δ Electorate between 1860 and 1866 and Δ eligible householders in 1866 in English boroughs along two axes. Panel D displays the values of Δ Electorate between 1869 and 1873 and Δ eligible householders in 1866 in English boroughs along two axes.

Figure A15: Democracy and Social Violence - ITT Estimates using Alternative Elections



(A) Elections 1865

(B) Elections 1868

(C) Elections 1874

NOTE: The first and second column of Panel B display the coefficients of the variable The variable *(log) Eligible Householders*$_{it}$ obtained using our main sample reported in columns 3 and 5 of Table 1, respectively. Panels A and C displays the results obtained using the period around the elections of 1865 [elections prior to the introduction of the Second Reform Act] and in 1874 [after the reform].

D.17 Placebo Events (Sports)

In this Appendix Section, we perform a further placebo exercise. Rather than using text algorithms to detect episodes of social violence, we apply the same tools to identify sport events. In other words, we use the sport-related variable to investigate whether the electoral reform affected sport events. The idea is that if the pattern found using sport-events is similar to the one obtained with social conflict-episodes, then we should be worried that the results displayed in Table 1 could be due to something else (for example, newspaper coverage or some hidden pattern in the data construction).

The results obtained using sport-related events in a borough (at the extensive margin) as dependent variable are displayed in Table A26. Corresponding estimates suggest no effect of the reform on sport events – which provides support to our empirical design.

Table A26: Placebo Analysis - Democracy and Sport Events

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Sport Events$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.00890	-0.0228	-0.0213				0.0197	-0.00252	-0.00355
	(0.0273)	(0.0286)	(0.0293)				(0.0366)	(0.0400)	(0.0397)
(log) Eligible Householders				0.0171	-0.00226	-0.00319			
				(0.0317)	(0.0358)	(0.0357)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.366	0.369	0.372	0.366	0.369	0.372			
1st stage F-Stat	-	-	-	-	-	-	232	194	184
City FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Gross Rental * Time FEs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Gini Gross Rental * Time FEs	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.217	.217	.217	.217	.217	.217	.217	.217	.217

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a sport event was observed in borough i and month t. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

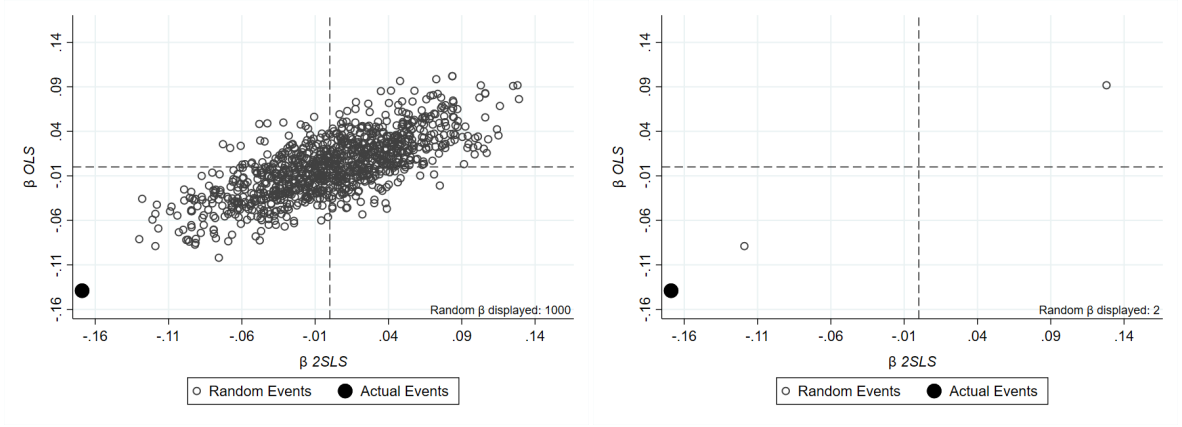
D.18 Placebo Samples

As discussed in Section 5.1 of the main text, to investigate concerns about our main findings having been obtained "by chance", we carry out a further placebo exercise where we randomly assign treatment in 1,000 placebo datasets with the same average social conflict likelihood as the "true" data (i.e. our main social conflict dataset built based on newspapers articles). Figure A16 below depicts the clouds of estimated coefficients of our baseline specifications (Columns 3 and 9 of baseline Table 1) with this "fake" data. Panel A displays all coefficients obtained from all 1,000 placebo samples. Each dot corresponds to one combination of coefficients in a Cartesian plane where the horizontal axis represents the OLS coefficient of the specification of Column 3, while the vertical axis depicts the 2SLS coefficient of the specification of Column 9. The large black dot represents our true coefficients. We can see that the cloud of placebo coefficients is centered around zero and it is extremely unlikely that the estimated coefficients of the baseline regressions could have been obtained "by chance".

In the same spirit, Panel B shows the estimates when the coefficients obtained with the two specifications (and the same placebo dataset) are both statistically significant at the 1 % level: This applies to only 2 placebo datasets (out of 1,000). These results highlight how extremely unlikely it would have been to obtain our results "by chance".

Further, we also display the results of an exercise where we have re-assigned randomly the treatment variable and the instrument at cross sectional level 1000 times. In Figure A17A [A17B] below, each dot corresponds to one combination of results in a Cartesian plane where the horizontal axis represents our beta coefficient of the OLS [2SLS] regressions, while the vertical axis depicts the corresponding p-value. The large black dot captures to the actual estimate in the regression, while the dashed line corresponds to a p-value=0.01. These findings highlight how unlikely it would have been to obtain similar results "by chance".

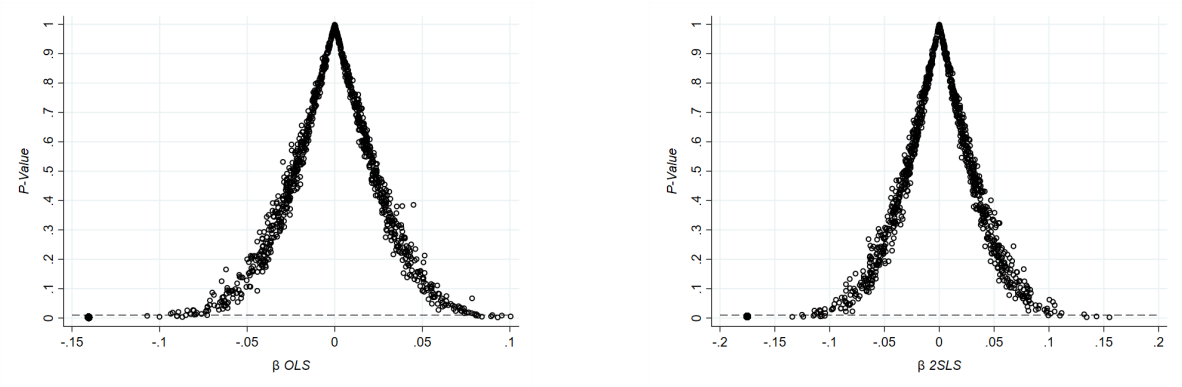
Figure A16: Results of Placebo Exercise



A) All estimated coefficients B) Both coefficients stat. sign. at the 1 % level

NOTE - Each panel displays all coefficients obtained using 1,000 placebo social conflict datasets with the same average social conflict likelihood as our main social conflict dataset built based on newspapers articles. Each dot corresponds to one combination of coefficients in a Cartesian plane where the horizontal axis represents the beta coefficient of the OLS specification of Column 3, while the vertical axis depicts the beta coefficient of the 2SLS specification of Column 9 of baseline Table 1. The large black dot represents our true coefficients. Panel A displays all coefficients. Panel B shows the estimates when the coefficients obtained with the two specifications (and the same placebo dataset) are both statistically significant at the 1 % level. The number of placebo datasets displayed in each cartesian plan is reported in the bottom-right corner.

Figure A17: Results of Placebo Exercise of Reshuffling Electors



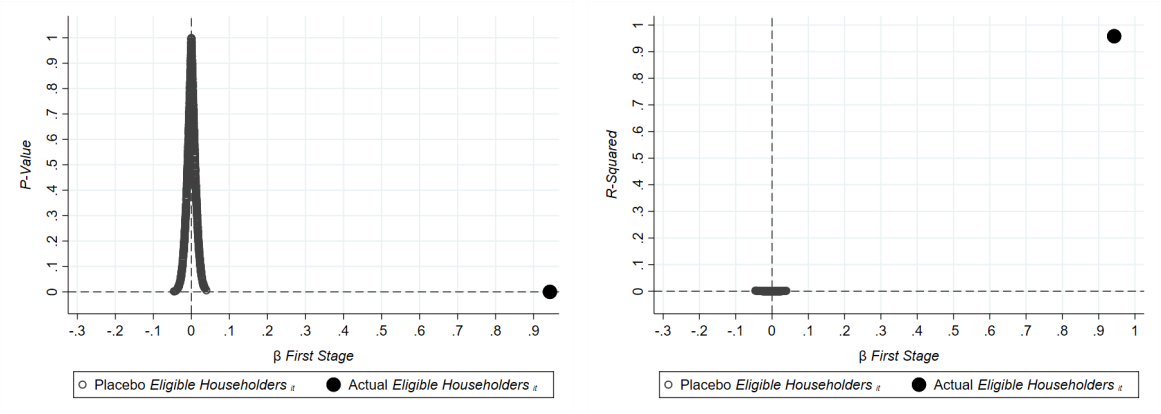
(A) Placebo Reshuffling Electors OLS (B) Placebo Reshuffling Electors 2SLS

NOTE: *Panel A* displays OLS coefficients estimated with our baseline regression (column 3 of Table 1) reshuffling the main independent variable *(Log) Electorate*$_{it}$ 1,000 times. *Panel B* displays 2SLS coefficients estimated reshuffling both the main independent variable *(Log) Electorate*$_{it}$ and *(Log) Eligible Electors*$_{it}$ 1,000 times (column 9 of Table 1). In *Panel A* the horizontal axis represents the beta coefficient of the OLS specification of Column 3 of Table 1, while the vertical axis depicts the corresponding p-value. In *Panel B* the horizontal axis represents the beta coefficient of the 2SLS specification of Column 9 of Table 1, while the vertical axis depicts the corresponding p-value. The large black dot represents our true coefficients.

D.19 Placebo Rent Structure

As mentioned in Section 5.1, to investigate concerns about the validity of our instrument, we carry out another placebo exercise where we randomly assign the structure of rents paid by households in 1,000 placebo datasets, with the distribution featuring the same average and standard deviation as the "true" data. Figure A18 below depicts the distribution of estimated first-stage coefficients with this "fake" rent data. In Panel A [Panel B], each dot corresponds to one combination of results in a Cartesian plane where the horizontal axis represents our first-stage beta coefficient, while the vertical axis depicts the corresponding p-value [r-squared value]. In both panels, the results obtained with the "true" rents distribution (represented by the large black dot) stick out from the cloud of results obtained with "fake" data. These findings highlight how unlikely it would have been to obtain similar first-stage results "by chance".

Figure A18: Placebo *(log) Eligible Householders*$_{it}$ and *(log) Electorate*$_{it}$



(A) $1st\ Stage\ \beta$ and $p-values$ (B) $1st\ Stage\ \beta$ and R^2

NOTE: Each panel displays all estimates obtained using 1,000 placebo datasets with the same average *(Log) Eligible Electors*$_{it}$ as the observed distribution. Each dot in Panel A [Panel B] corresponds to one combination of results in a Cartesian plane where the horizontal axis represents our first-stage beta coefficient, while the vertical axis depicts the corresponding p-value [r-squared value]. The large dots display the results obtained with true values of *(Log) Eligible Electors*$_{it}$ observed in our data.

D.20 Alternative Data Construction Methods

In what follows we investigate the robustness of our findings with respect to technical details of our social conflict variable construction. The baseline results were obtained by extracting boroughs in the string [-25:+50] nearby the social conflict related words. In the current Appendix Section we explore whether our results hold when we adopt different string bandwidths. As shown in Tables A27 to A29, the baseline findings are not sensitive to the string bandwidths used.

Table A27: Democracy and Social Violence - Alternative String Bandwidths [1]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.127***	-0.126***	-0.128***				-0.152***	-0.158***	-0.156***
	(0.0412)	(0.0446)	(0.0457)				(0.0527)	(0.0599)	(0.0590)
(log) Eligible Householders$_{it}$				-0.132***	-0.141***	-0.141***			
				(0.0454)	(0.0529)	(0.0525)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.383	0.387	0.392	0.383	0.386	0.391	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.202	.202	.202	.202	.202	.202	.202	.202	.202

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.20 by extracting boroughs in the string [0:+50] nearby social conflict related words. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A28: Democracy and Social Violence - Alternative String Bandwidths [2]

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.146***	-0.152***	-0.153***				-0.155***	-0.171***	-0.170***
	(0.0380)	(0.0411)	(0.0421)				(0.0493)	(0.0557)	(0.0551)
(log) Eligible Householders$_{it}$				-0.135***	-0.153***	-0.153***			
				(0.0426)	(0.0489)	(0.0489)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.397	0.399	0.404	0.395	0.398	0.403	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.298	.298	.298	.298	.298	.298	.298	.298	.298

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.20 by extracting boroughs in the string [-50:+50] nearby social conflict related words.The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A29: Democracy and Social Violence - Alternative String Bandwidths [3]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.128***	-0.129***	-0.130***				-0.144**	-0.150**	-0.149**
	(0.0429)	(0.0452)	(0.0464)				(0.0567)	(0.0620)	(0.0615)
(log) Eligible Householders$_{it}$				-0.125**	-0.134**	-0.134**			
				(0.0487)	(0.0547)	(0.0546)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.400	0.402	0.408	0.399	0.401	0.407	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.303	.303	.303	.303	.303	.303	.303	.303	.303

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.20 by extracting boroughs in the string [-25:+75] nearby social conflict related words. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.21 Social Conflict-related Keywords

An important parameter for the construction of the social conflict data is the set of keywords used, as discussed in Section 3. In Table A30 below are listed the baseline set of terms used to identify social conflict-related sentences. In the current Appendix Section we assess whether our results are robust when only a subsample of these keywords are used. We first replicate our baseline results when removing from the variable construction all *disturbance*-related keywords (i.e., disturbance, disturbances). Corresponding results are displayed in Table A31. We find in all specifications a strong and statistically significant impact of the explanatory variable on reducing social conflict.

Table A30: List of Social Conflict-related Keywords

Disturbance, Disturbances
Unrest
Riot, Riots, Rioters, Rioting
Tumult, Tumults
Disorder, Disorders

Similar results are obtained when we remove social conflict events identified with the keyword *unrest* (Table A32), *riot*-related keywords (Table A33), *disorder*-related keywords (Table A34) and *tumult*-related keywords (Table A35). In all cases, the baseline results prove robust to modifying the set of keywords included.

Finally, we also include a robustness table where we add to the set of keywords the additional keywords "protest, protests and protesters". These keywords are somewhat more frequent than others, which may entail the risk of picking up events that are not clearly related to social conflict. Still it is useful to study the sensitivity of findings with respect to this extended social conflict keyword definition. The corresponding results are reported in Table A36 and are in line with the baseline results presented in the main text.

Table A31: Democracy and Social Violence - Exclude *Disturbance*-related Events

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.112***	-0.110***	-0.113**				-0.155***	-0.163***	-0.162***
	(0.0408)	(0.0421)	(0.0435)				(0.0522)	(0.0565)	(0.0557)
(log) Eligible Householders$_{it}$				-0.134***	-0.146***	-0.146***			
				(0.0444)	(0.0490)	(0.0485)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.399	0.403	0.406	0.399	0.404	0.406	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ ** Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.253	.253	.253	.253	.253	.253	.253	.253	.253

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1 using the following keywords: Unrest, Riot, Riots, Rioters, Rioting, Tumult, Tumults, Disorder, Disorders.The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A32: Democracy and Social Violence - Exclude *Unrest*-related Events

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.132***	-0.141***	-0.143***				-0.170***	-0.199***	-0.198***
	(0.0436)	(0.0459)	(0.0474)				(0.0577)	(0.0639)	(0.0633)
(log) Eligible Householders$_{it}$				-0.147***	-0.178***	-0.178***			
				(0.0495)	(0.0555)	(0.0552)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.398	0.402	0.405	0.398	0.402	0.405	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ * *Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.291	.291	.291	.291	.291	.291	.291	.291	.291

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1 using the following keywords: Disturbance, Disturbances, Riot, Riots, Rioters, Rioting, Tumult, Tumults, Disorder, Disorders. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A33: Democracy and Social Violence - Exclude *Riot*-related Events

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.0828**	-0.0994**	-0.0997**				-0.0820	-0.118**	-0.118**
	(0.0384)	(0.0399)	(0.0407)				(0.0543)	(0.0574)	(0.0570)
(log) Eligible Householders$_{it}$				-0.0711	-0.106**	-0.106**			
				(0.0475)	(0.0515)	(0.0514)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.375	0.378	0.382	0.375	0.378	0.382	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ ** Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.227	.227	.227	.227	.227	.227	.227	.227	.227

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1 using the following keywords: Disturbance, Disturbances, Unrest, Tumult, Tumults, Disorder, Disorders. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A34: Democracy and Social Violence - Exclude *Disorder*-related Events

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.139***	-0.140***	-0.141***				-0.183***	-0.198***	-0.197***
	(0.0438)	(0.0458)	(0.0472)				(0.0582)	(0.0644)	(0.0637)
(log) Eligible Householders$_{it}$				-0.158***	-0.178***	-0.178***			
				(0.0498)	(0.0560)	(0.0557)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.398	0.401	0.404	0.398	0.401	0.404	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ * *Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.278	.278	.278	.278	.278	.278	.278	.278	.278

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1 using the following keywords: Disturbance, Disturbances, Unrest, Riot, Riots, Rioters, Rioting, Tumult, Tumults.The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A35: Democracy and Social Violence - Exclude *Tumult*-related Events

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.138***	-0.145***	-0.147***				-0.174***	-0.201***	-0.200***
	(0.0432)	(0.0455)	(0.0470)				(0.0571)	(0.0636)	(0.0629)
(log) Eligible Householders$_{it}$				-0.151***	-0.180***	-0.180***			
				(0.0490)	(0.0552)	(0.0549)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.400	0.404	0.406	0.400	0.404	0.406	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ * *Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.288	.288	.288	.288	.288	.288	.288	.288	.288

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1 using the following keywords: Disturbance, Disturbances, Unrest, Riot, Riots, Rioters, Rioting, Disorder, Disorders. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A36: Democracy and Social Conflict - Adding *Protest*-Related Events

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.137***	-0.143***	-0.145***				-0.177***	-0.202***	-0.200***
	(0.0436)	(0.0459)	(0.0474)				(0.0575)	(0.0640)	(0.0634)
(log) Eligible Householders$_{it}$				-0.154***	-0.181***	-0.180***			
				(0.0492)	(0.0555)	(0.0552)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.401	0.404	0.407	0.401	0.405	0.408	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ ** Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.292	.292	.292	.292	.292	.292	.292	.292	.292

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1, and using all baseline conflict-related keywords plus an additional set of protest-related keywords. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.22 Building the Social Conflict Variable Using a Machine Learning (Lasso) Approach

As mentioned in Section 5.1, our main social conflict variable is constructed using a bag-of-words method. This has the virtue of simplicity and transparency but may be somewhat sensitive to the keywords used. While in the previous Appendix Section we have already performed a first robustness check in that respect, the current Appendix Section goes one step further and completely re-creates from scratch an alternative social conflict measure that relies on a very different approach.

In particular, what follows makes use of machine learning (i.e. a lasso model) to build our social conflict measure. We start with 1,000 sentences on violent events identified by our baseline bag-of-words algorithm. Then we check manually each entry and code whether it indeed refers to a violent event in the corresponding borough or whether it is a "false positive". Out of 1,000 sentences, around 85 % of our strings are related to true violent events. We use 900 strings to train an algorithm (i.e. cross-repetition lasso using keyword in the string as explanatory variables) that predicts if the sentence is related to a true social conflict or not. With our algorithm we are able to reach a goodness of prediction of 93% in 100 out-of-sample articles. In particular, we train the lasso model on 900 articles, and using the trained model, we are able to identify correctly [i.e., string with event or string without event] 93 out of 100 out-of-sample strings that were not included in the training.[34]

We then apply our algorithm to the full set of strings, and only keep the ones identified by the algorithm as social conflict-related to construct the corresponding dependent variable. Note that the sample mean of social conflict events obtained using this alternative approach is very similar to our baseline. This is due to the fact that our dependent variable of interest is a dummy (e.g., if there are two strings in a month coded as social conflict-related by the original algorithm, the probability of having no true event is only of 2.25%).

Using this alternative lasso-based social conflict measure, we estimate a series of robustness specifications (Tables A37 to A39). In all cases we find very similar results to our baseline estimates.

We also explore whether errors (i.e., when lasso events are equal to 0 and original events are equal to 1) are correlated with our identifying source of variation (i.e., whether errors are more likely in places that enfranchise more under the reform) but this does not appear to be the case, as shown in Table A40. In this table, the dependent variable is equal to 1 if there is at least one social conflict event in the borough using all strings identified by the baseline

[34] We also explored whether using a larger string increases the accuracy of the prediction. It turns out that goodness of prediction is not affected by the length of the string (i.e., short strings contain enough information to detect social violence events). In this second exercise we use strings with average length of 368 (median 400) [around 4 times larger than the *baseline* string]. Using this second set of data, out-of-sample accuracy is similar (94%).

bag-of-words approach, but no events found in the borough when applying the lasso method. We can see that this measure of "false positives" is unrelated to our key political explanatory variables of interest. These results are consistent with the notion that measurement error may be of a "classical" type, leading –if anything– to attenuation bias.

Table A37: Democracy and Social Violence - Using Results from Lasso-exercise [1]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence [LASSO]$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.140***	-0.141***	-0.143***				-0.169***	-0.180***	-0.179***
	(0.0420)	(0.0445)	(0.0456)				(0.0561)	(0.0626)	(0.0620)
(log) Eligible Householders$_{it}$				-0.146***	-0.162***	-0.161***			
				(0.0482)	(0.0549)	(0.0547)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.374	0.377	0.382	0.373	0.376	0.381	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.23	.23	.23	.23	.23	.23	.23	.23	.23

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.22. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A38: Democracy and Social Violence - Using Results from Lasso-exercise [2]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: Social Violence [LASSO]$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	-0.196***	-0.204***	-0.207***				-0.233**	-0.259**	-0.257***
	(0.0651)	(0.0674)	(0.0700)				(0.0912)	(0.1000)	(0.0977)
(log) Eligible Householders$_{it}$				-0.202**	-0.232***	-0.232***			
				(0.0790)	(0.0885)	(0.0873)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.519	0.522	0.527	0.518	0.521	0.526	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.263	.263	.263	.263	.263	.263	.263	.263	.263

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is the (log+1) number of violent events observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.22. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A39: Democracy and Social Violence - Using Results from Lasso-exercise [3]

Dep. Variable: IHS Social Violence [LASSO]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.248***	-0.258***	-0.262***				-0.293**	-0.327***	-0.324***
	(0.0817)	(0.0847)	(0.0878)				(0.114)	(0.125)	(0.122)
(log) Eligible Householders$_{it}$				-0.254**	-0.293***	-0.292***			
				(0.0986)	(0.110)	(0.109)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.517	0.519	0.525	0.516	0.519	0.524	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.337	.337	.337	.337	.337	.337	.337	.337	.337

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is the inverse hyperbolic sine transformation of violent events observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.22. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A40: Difference between Lasso-events and Baseline-events

Dep. Variable: Social Violence ≠ Social Violence [LASSO]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	0.000937	0.00271	0.00296				0.000285	0.00397	0.00381
	(0.00282)	(0.00267)	(0.00288)				(0.00363)	(0.00275)	(0.00276)
(log) Eligible Householders$_{it}$				0.000247	0.00356	0.00343			
				(0.00315)	(0.00244)	(0.00244)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.787	0.792	0.793	0.786	0.792	0.793			
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Gross Rental * Month FEs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Gini Gross Rental * Month FEs	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.007	.007	.007	.007	.007	.007	.007	.007	.007

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is equal to 1 if there is at least one social conflict event in the borough i and month t using all strings identified by the baseline bag-of-words approach, but no events found in the borough when applying the lasso method. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.22. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.23 Building the Social Conflict Variable Using Natural Language Processing

An alternative approach to our bag-of-words method is to rely on natural language processing (NLP) (see e.g. Fetzer, 2020). In what follows we perform a robustness exercise where we exploit an NLP alternative algorithm to detect social conflict events. Before discussing our findings with this algorithm, we highlight some technical challenges that are worth noting for our application in the next paragraphs.

First, we are unable to access the full set of articles. The British Newspaper Archive is the largest repository of digitized archives of British and Irish newspapers. There are over 46 millions pages available on the website (and several millions for our sample period alone). This massive number prevents us from downloading the entire corpus and we are forced to select a limited number of conflict related words and then download only articles that contain these.

Second, natural language processing algorithms crucially rely on the quality of the raw text. While articles that are obtained from digital sources (i.e., newspaper articles available online) are close to perfect, when using historical data we face a problem related to the quality of the Optical Character Recognition process [OCR] (i.e., which converts the image of the article into text). Given that with historical data the use of OCR implies various typos, having too sophisticated methods relying on the exact sentence structure may be outperformed by simpler yet more robust bag-of-words approaches.

Third, for historical newspaper articles the low performance of OCR quality manifests itself also in the process of sentence boundary detection (i.e., the process of splitting a text into sentences).

Finally, the NLP algorithm is several times more computationally taxing than our baseline method.

Despite the aforementioned caveats we have used a Natural Language Processing algorithm (i.e. we use SPACY, an open-source software library that uses statistical neural network models to perform computational linguistic tasks) to identify entities.[35] From an operational point of view, the procedure to build a conflict dataset using this alternative algorithm consists of three sequential steps. Firstly, we take the full text of the article and we identify all sentences with conflict-related keywords. Then, we perform the named entity recognition to extract locations from the text. Finally, we match locations with the boroughs in our dataset. The corresponding findings for this alternative algorithm are presented in Table A41. The results are in line with the baseline estimates of the main text.

[35] Here is a demo of the built-in named entity pre-trained model used by SPACY.

Table A41: Democracy and Social Violence - Using SPACY Sentencizer

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.105***	-0.103**	-0.104**				-0.140***	-0.147**	-0.147**
	(0.0395)	(0.0425)	(0.0432)				(0.0526)	(0.0586)	(0.0580)
(log) Eligible Householders$_{it}$				-0.122***	-0.132**	-0.132**			
				(0.0446)	(0.0514)	(0.0515)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.400	0.402	0.407	0.400	0.402	0.407	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.27	.27	.27	.27	.27	.27	.27	.27	.27

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section D.23. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.24 Restricting to social conflict events of inter-regional and national importance

One concern could be that our algorithm may not exclusively capture important instances of political violence, but may also occasionally pick up some purely local events (e.g. bar rows). In order to address this worry, in this appendix section we construct a variant of our social violence measure that only takes into account events that are important enough that they are also reported from newspapers with headquarters in other boroughs. For example, if a given event taking place in Liverpool is only reported in Liverpool-based newspapers it is discarded, while if it is also reported in newspapers located in other boroughs it is kept in the sample. This restriction results in a focus on typically bigger events of inter-regional and national scope.

Table A42 below reports the findings. As benchmark, column 1 depicts the coefficients obtained using the full sample of newspapers available. In contrast, the results displayed in column 2 are obtained using only events reported in newspapers published in a location that is at least 50 km away from borough i. As shown in this column, the mean social conflict risk is (mechanically) somewhat smaller when restricting ourselves to such bigger events, but the estimated coefficients remain of similar magnitude. This implies that our results are – if anything– stronger when focusing on bigger events. Similarly, in column 3 the minimum distance imposed between newspaper location and borough i is 100 km, and so on, up to a distance restriction of 350 km in column 8. Throughout Table A42 we find evidence for a strong impact of enfranchisement on social violence, even when restricting our attention to large-scale events of national importance covered by far-away media.

Table A42: Democracy and Social Violence - Using newspapers with a certain distance from borough i

Dep. Variable: Social Violence$_{it}$	-	*>50 km*	*>100 km*	*>150 km*	*>200 km*	*> 250 km*	*>300 km*	*>350 km*
				Panel A: OLS Results				
(log) Electorate$_{it}$	-0.141***	-0.117**	-0.120***	-0.116***	-0.110***	-0.0957**	-0.0699**	-0.0694**
	(0.0457)	(0.0455)	(0.0415)	(0.0383)	(0.0383)	(0.0368)	(0.0343)	(0.0313)
				Panel B: Reduced Form Results				
(log) Eligible Householders$_{it}$	-0.158***	-0.165***	-0.152***	-0.148***	-0.147***	-0.117***	-0.0933**	-0.101***
	(0.0549)	(0.0528)	(0.0477)	(0.0451)	(0.0449)	(0.0445)	(0.0415)	(0.0381)
				Panel C: Two-Stage Least Square Results				
(log) Electorate$_{it}$	-0.175***	-0.184***	-0.168***	-0.165***	-0.163***	-0.130**	-0.104**	-0.112**
	(0.0620)	(0.0608)	(0.0547)	(0.0522)	(0.0522)	(0.0505)	(0.0471)	(0.0438)
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	.237	.192	.172	.154	.139	.125	.094	.078

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM [2SLS] estimates are reported in Panels A and B [Panel C]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. Results displayed in column 1 are obtained using the full sample of newspapers available. Results reported in column 2 are obtained using only events reported in newspapers published in a location that is at least 50 km away from borough i, in column 3 the minimum distance between newspaper location and borough i is 100 km, and so on. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). All columns include Borough FEs, Time FEs and flexible controls of average and Gini rent variables (i.e., we allow the linear effect of rent-based controls to vary at the month-year level). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E. Appendix: Channels

E.1 Voice: Non-linear Effect of Enfranchisement

In this Appendix we provide additional evidence of the non-linear effect of enfranchisement on voter participation.

To do so, we divide the boroughs into 4 groups based on the intensity of enfranchisement and we explore whether we observe a differential increase in the numbers of voters across different groups in the 1868 elections.

Table A43 displays the corresponding results. The explanatory variables of interest are the interactions between Δ *Electorate* - X^X *Quartile*$_{it}$ and *Elections 1868*. The first component of this interaction corresponds to a dummy that takes a value of 1 if the increase of electorate in borough i is in the X^X quartile, while the second component is a dummy *Elections 1868*$_t$ that takes a value of 1 for the election of 1868.

The corresponding results show that boroughs that experienced an increase above the bottom quartile (the reference category) feature higher numbers of voters and this increase is larger, the greater is the intensity of enfranchisement. Similar results are reported in Columns 4 to 6 when using explanatory variables based on the number of Eligible Householders.

Table A43: Democracy and Participation in the Political Arena - Additional Results

Dep. Var: (log) Voters$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF
Δ Electorate - 2nd Quartile * Elections 1868	0.282**	0.223	0.170			
	(0.131)	(0.141)	(0.138)			
Δ Electorate - 3rd Quartile * Elections 1868	0.600***	0.531***	0.480***			
	(0.0929)	(0.0903)	(0.0715)			
Δ Electorate - 4th Quartile * Elections 1868	0.966***	0.908***	0.861***			
	(0.101)	(0.0966)	(0.0855)			
Δ Eligible Householders - 2nd Quartile * Elections 1868				0.405***	0.345**	0.298**
				(0.133)	(0.133)	(0.114)
Δ Eligible Householders - 3rd Quartile * Elections 1868				0.481***	0.425***	0.341***
				(0.141)	(0.142)	(0.121)
Δ Eligible Householders - 4th Quartile * Elections 1868				0.705***	0.638***	0.674***
				(0.156)	(0.157)	(0.156)
Observations	216	216	216	216	216	216
R-squared	0.972	0.973	0.974	0.961	0.962	0.965
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Gross Rental * Time FEs	No	Yes	Yes	No	Yes	Yes
Gini Gross Rental * Time FEs	No	No	Yes	No	No	Yes

Notes: The unit of observation is a borough i and election t (where t corresponds to the elections of 1865 and 1868). OLS estimates are reported in all columns. The dependent variable is the *(log) Voters$_{it}$* correspond to the electorate in a borough i observed in election t. The variable Δ *Electorate - X^X Quartile$_{it}$* is a dummy that takes a value of 1 if the increase of electorate in borough i is in the X^X quartile. The variable Δ *Electorate$_i$* is computed as $log((Electors\ Post - Reform)/(Electors\ Pre - Reform))_i$ where $(Electors\ Post - Reform)_i$ and $(Electors\ Pre - Reform)_i$ correspond to the electorate pre and after the Second Reform Act in a city i, respectively. The dummy *Elections 1868$_t$* takes a value of 1 for periuod corresponding to the election of 1868. The variable Δ *Eligible Householders - X^X Quartile$_{it}$* is a dummy that takes a value of 1 if the increase of eligible householders in borough i is in the X^X quartile. The variable Δ *Eligible Householders$_i$* is computed as $log((Number\ Householders\ 4 - 10£ + Number\ Householders\ Above\ 10£)/(Number\ Householders\ Above\ 10£))_i$, where $(Number\ Householders\ 4 - 10£)_i$ and $(Number\ Householders\ Above\ 10£)_i$ correspond to the number of householders living in a house with rental value between 4 and 10 pounds (who were previously were banned from voting but enfranchised after the Second Reform Act) and the number of householders with rental value above 10 pounds (who could already vote before the reform) in a borough i, respectively. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E.2 Voice: Increased Participation in the Political Arena and Beyond

In this Appendix section we display further results on the impact of enfranchisement on overall political engagement and competitiveness.

Voter Turnout As mentioned in the main text, we start by investigating the impact of franchise extension on voter turnout. To study this, we run the following regression:

$$(log)\ Turnout_{i,t} = \beta_0 + \beta_1 (log)\ Electorate_{i,t} + \epsilon_{it} \tag{3}$$

Table A44: Channels - Democracy and (log) Turnout

Dep. Variable: (log) Turnout$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	0.102	0.0734	0.0425				0.226	0.197	0.180
	(0.105)	(0.0938)	(0.0722)				(0.194)	(0.193)	(0.178)
(log) Eligible Householders$_{it}$				0.157	0.140	0.126			
				(0.134)	(0.135)	(0.123)			
Observations	216	216	216	216	216	216	216	216	216
R-squared	0.574	0.576	0.589	0.580	0.581	0.594	-	-	-
1st stage F-Stat Delta Electors	-	-	-	-	-	-	87	65	57
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes

Notes: The unit of observation is borough i and election t (where t corresponds to the elections of 1865 or 1868).The sample covers 108 boroughs. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is the *(log) Turnout$_{it}$* computed as the ratio of voters over total electorate in a borough i for the elections of 1865 and 1868, respectively.The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A44 is imprecisely estimated: The coefficient of the explanatory variable of interest (change in electorate) is always positive but not statistically significant.

Plurality Moreover, as mentioned in the main text, here we also report the results when we focus on boroughs in which the newly enfranchised people constitute a relative majority of the total number of voters and consequently have more political clout. In Table A45 below we regress our main dependent variable of social conflict on two different plurality dummies interacted with the post-election period. In particular, in the odd columns the variable $Plurality_i$ takes a value of 1 if the newly enfranchised voters constitute a majority of the total

number of voters in a town and 0 otherwise. In the even columns the dummy $Plurality66_i$ takes a value of 1 if the number of new voters amounts to more than 2/3 (66%) of the total number of voters (a qualified (super-)majority). Analogous to our main baseline specification, we instrument these dummies using the number of householders in the appropriate rent bands. For example, the dummy $Plurality_i$ is instrumented with a dummy that takes a value of 1 if the number of householders between $£4$ and $£10$ (newly eligible householders) is greater than the number of householders above $£10$ (previously eligible householders). Table A45 shows that new voters enhancing plurality significantly decreases social violence and that the more newly enfranchised voters there are (and hence the more "voice" they have), the more social conflicts decrease.

Table A45: Democracy and Plurality - New Voters as Relative and Qualified Majority

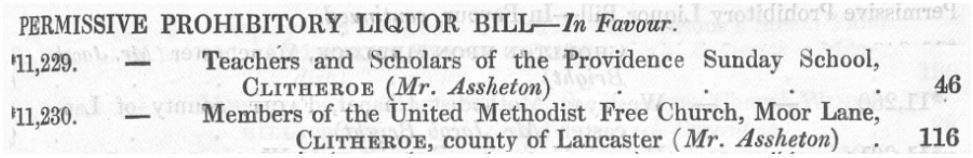
Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) RF	(4) RF	(5) 2SLS	(6) 2SLS
Plurality$_i$ * *Post - Elections 1868*$_t$	-0.0935***				-0.106*	
	(0.0292)				(0.0603)	
Plurality66$_i$ * *Post - Elections 1868*$_t$		-0.0946***				-0.233***
		(0.0329)				(0.0745)
Plurality Eligible Householders$_i$ * *Post - Elections 1868*$_t$			-0.0491*			
			(0.0291)			
Plurality Eligible Householders66$_i$ * *Post - Elections 1868*$_t$				-0.139***		
				(0.0429)		
Observations	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.395	0.395	0.393	0.396	-	-
1st stage F-Stat	-	-	-	-	38	52
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes
Rent Inequality$_i$ * *Time FEs*	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-4 [5-6]. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *Plurality*$_i$ * *Post - Elections 1868*$_t$ corresponds to the interaction between two indicator variables: *Plurality*$_i$ which takes a value of 1 if the newly enfranchised voters constitute a majority of the total number of voters (and 0 otherwise) and *Post - Elections 1868*$_t$ which takes value 1 after the Elections of 1868. The variable *Plurality Eligible Householders*$_i$ * *Post - Elections 1868*$_t$ corresponds to the interaction between two indicator variables: *Plurality Eligible Householders*$_i$ takes a value of 1 if the number of householders between 4£and 10£(newly eligible householders) is greater than the number of householders above 10£(previously eligible householders) and *Post - Elections 1868*$_t$ which takes value 1 after the Elections of 1868. The variables *Plurality66*$_i$ * *Post - Elections 1868*$_t$ and *Plurality Eligible Householders66*$_i$ * *Post - Elections 1868*$_t$ are analogously defined but for a qualified majority of 66% instead of simple majority of 50%. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Petitions In what follows, we display supporting material for the results on petitions. We start with a display of examples of petitions in Figures A19A and A19B. Next, Figure A20 depicts the distribution of petitions, which suggests the use of a Poisson estimator. Finally,

Table A46 shows that the baseline result on petitions from the main text extend to a longer sample period (for which a greater number of observations are available).

Figure A19: Examples of petitions

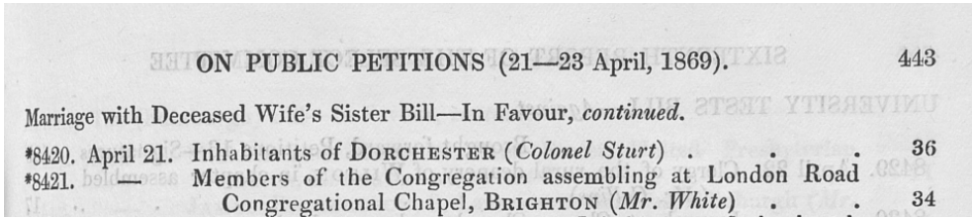


PERMISSIVE PROHIBITORY LIQUOR BILL—*In Favour.*

11,229. — Teachers and Scholars of the Providence Sunday School, Clitheroe (*Mr. Assheton*) 46

11,230. — Members of the United Methodist Free Church, Moor Lane, Clitheroe, county of Lancaster (*Mr. Assheton*) . 116

(A) Petitions from Clitheroe in *May 1869*



ON PUBLIC PETITIONS (21—23 April, 1869). 443

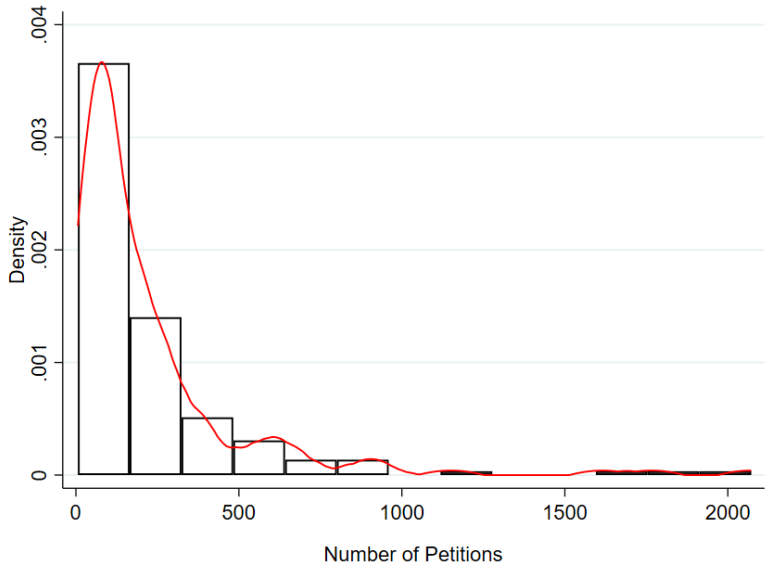
Marriage with Deceased Wife's Sister Bill—In Favour, *continued.*

*8420. April 21. Inhabitants of Dorchester (*Colonel Sturt*) . . . 36

*8421. — Members of the Congregation assembling at London Road Congregational Chapel, Brighton (*Mr. White*) . . 34

(B) Petition from Dorchester in *April 1869*

Figure A20: Distribution of petitions received at the borough-level over the period



NOTE: The graph shows the distribution of petitions observed from 1868 to 1869 at the borough-level. Data on petitions is based on the archival records of the Select Committee on Public Petitions that contains all public petitions received by the House of Commons over the period.

Table A46: Democracy and Petitions - Extended Sample

Dep. Variable: Number of petitions submitted$_{it}$	(1)	(2)	(3)	(4)
(log) Electorate$_{it}$	11.25***	0.420*		
	(2.884)	(0.217)		
(log) Eligible Householders			10.03***	0.654***
			(1.439)	(0.238)
Observations	1,840	6,992	1,840	6,992
Controls	Yes	Yes	Yes	Yes
Town-Linear Trend	No	Yes	No	Yes
Sample Period	Baseline	1866-1871	Baseline	1866-1871

Notes: The unit of observation is borough i and month t. Poisson estimates are reported in all columns. The sample used in columns 1 and 3 (2 and 4) covers the period 1868-1869 (1866-1871). The dependent variable is the number of petitions submitted in borough i and month t. The petition data was constructed using archival records of the Selected Committee on Public Petitions (Parliament (n.d.)). The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). All columns include Borough FEs, Time FEs and flexible controls of average and Gini rent variables (i.e., we allow the linear effect of rent-based controls to vary at the month-year level). Columns (2) and (4) include borough-specific linear trend. Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E.3 Increase in Economic Activity: Further Results

In this Appendix Section we carry out further investigations of the "enfranchisement boosting growth" channel. First, in Appendix Table A47 we replicate Table 5, but relying on Poisson estimations instead of OLS, which yields very similar results.

Second, given that our proxy for economic activity is the number of job advertisements mentioning a borough, could it be that our results are driven by composition effects? As found by Fergusson, Riano and Song (2020), media penetration may contribute to the success of institutional reforms, and hence it could be that our results on job advertisements may be due to boroughs with a greater franchise extension happening to enjoy greater media penetration and simply getting more mentions in general, not specifically linked to job advertisements. As shown in the "placebo" Table A48, franchise extension does *not* predict general mentions of a given borough in the media (i.e. the (log) number of articles available in the *British Newspaper Archive* mentioning the borough i in month t). This attenuates concerns about our results being driven by underlying media penetration.

Finally, in Table A49 we investigate whether there may indeed be a causal relationship of economic growth affecting political stability. In particular, we regress the dependent variable of social conflict on the *predicted* level of job advertisements in a given borough.[36] In particular, the variable *[Pred.] (log) borough Mentions in Newspaper Ads*$_{it}$ represents the average predicted value of *Borough Mentions in Newspaper Ads* obtained using the specification of Table 5, column 1 and excluding from the estimation sample borough i itself. For each borough i, the variable *[Pred.] (log) borough Mentions in Newspaper Ads*$_{it}$ is computed using all other boroughs with a similar level of Δ $Electorate$ [+/- 10%].[37] As shown in Table A49, we indeed find that the predicted job openings are a strong predictor of a decline in social conflicts.

[36] As there may be a bi-directional relationship between social conflict and growth, to focus on the link between economic growth and social conflict, and for the sake of addressing endogeneity concerns, it is preferable to focus on the *predicted* rather than *actual* level of job advertisements.

[37] One borough (Merthyr Tydfil) drops from the sample, as there was no borough with a similar (enough) delta electorate.

Table A47: Channels - Democracy and Economic Growth - Newspapers Ads (Poisson)

Dep. Variable: Borough Mentions in Newsp. Ads$_{it}$	(1)	(2)	(3)	(4)
(log) Electorate$_{it}$	0.106***	0.108***		
	(0.0295)	(0.0375)		
(log) Eligible Householders$_{it}$			0.119***	0.136***
			(0.0278)	(0.0353)
Observations	4,416	4,416	4,416	4,416
Borough FEs	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	Yes	No
Rent Inequality$_i$ ** Time FEs*	No	No	Yes	No
Sample Mean	122.595	122.595	122.595	122.595

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. Poisson estimates are reported in columns 1-6. The dependent variable is the number of mentions of borough i in pages of job advertisements in month t using national or local newspapers available on the *British Newspaper Archive*. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A48: Placebo - Enfranchisement and Borough Mentions in News Articles

Dep. Variable: (log) Borough Mentions in Newsp. Articles$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.00397	-0.0101	-0.0104				-0.0377	-0.0608	-0.0604
	(0.0406)	(0.0444)	(0.0448)				(0.0481)	(0.0565)	(0.0565)
(log) Eligible Householders$_{it}$				-0.0327	-0.0545	-0.0544			
				(0.0417)	(0.0507)	(0.0510)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.957	0.957	0.959	0.957	0.957	0.959	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*$_t$	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ * *Time FEs*$_t$	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	6.904	6.904	6.904	6.904	6.904	6.904	6.904	6.904	6.904

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is (log) number of articles available on the *British Newspaper Archive* mentioning the borough i in month t. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A49: Predicted Economic Growth and Social Violence

Dep. Variable: Social Violence$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
[Pred.] (log) Borough Mentions in News. Ads$_{it}$	-1.122***	-1.127***						
	(0.296)	(0.310)						
[Pred.] (log) Borough Mentions in News. Ads [> 50 km]$_{it}$			-1.094***	-1.094***				
			(0.279)	(0.289)				
[Pred.] (log) Borough Mentions in News. Ads [> 75 km]$_{it}$					-1.067***	-1.064***		
					(0.278)	(0.287)		
[Pred.] (log) Borough Mentions in News. Ads [> 100 km]$_{it}$							-1.057***	-1.050***
							(0.276)	(0.283)
Observations	4,392	4,392	4,392	4,392	4,392	4,392	4,392	4,392
R-squared	0.390	0.398	0.391	0.398	0.391	0.398	0.391	0.398
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes	No	Yes
Rent Inequality$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *[Pred.] (log) borough Mentions in Newspaper Ads*$_{it}$ represents the average predicted value of *Borough Mentions in Newspaper Ads* obtained using the specification of Table 5, column 1 and excluding from the estimation sample borough i itself. For each borough i, the variable *[Pred.] (log) borough Mentions in Newspaper Ads*$_{it}$ is computed using all other boroughs with similar level of Δ $Electorate$ [+/- 10%]. The variable *[Pred.] (log) borough Mentions in Newspaper Ads [>50 km]*$_{it}$ [>75 km] (>100 km) is computed as the average of predicted values of all other boroughs with similar level of Δ $Electorate$ [+/- 10%] and with a distance of at least 50 [75] (100) km from borough i. Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E.4 Increase in Economic Activity: Alternative Measures of Newspaper-Based Index of Economic Activity

In this Appendix Section we study alternative definitions of our newspaper-based index of economic activity. In Tables A50 and A51 we replicate Table 5 using the (log) total number of ads containing the name of the borough and the difference of in (log) of number of ads containing the word "Wanted" and (log) total number of articles containing the name of the borough, respectively. The corresponding results are similar to the estimates presented in the main text, hence providing further evidence of the positive effect of enfranchisement on boosting growth.

Table A50: Democracy and Economic Growth - Alternative Dependent Variable [1]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable: (log) Borough Mentions in Newsp. Ads [ALL]$_{it}$	OLS	OLS	OLS	RF	RF	RF	2SLS	2SLS	2SLS
(log) Electorate$_{it}$	0.166***	0.178***	0.189***				0.220***	0.262***	0.255***
	(0.0445)	(0.0493)	(0.0474)				(0.0524)	(0.0622)	(0.0605)
(log) Eligible Householders$_{it}$				0.191***	0.235***	0.230***			
				(0.0480)	(0.0575)	(0.0557)			
Observations	4,415	4,415	4,415	4,415	4,415	4,415	4,415	4,415	4,415
R-squared	0.961	0.961	0.962	0.961	0.961	0.962	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	4.911	4.911	4.911	4.911	4.911	4.911	4.911	4.911	4.911

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is (log+1) of number of mentions of borough i in pages of job advertisements in month t using national or local newspapers available on the *British Newspaper Archive*. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A51: Democracy and Economic Growth - Alternative Dependent Variable [2]

Dep. Variable: (log) City Mentions in Newsp. Ads - (log) Borough Mentions in Articles$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	0.179***	0.177***	0.196***				0.286***	0.313***	0.301***
	(0.0614)	(0.0668)	(0.0609)				(0.0718)	(0.0788)	(0.0750)
(log) Eligible Householders$_{it}$				0.248***	0.281***	0.271***			
				(0.0640)	(0.0721)	(0.0680)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.687	0.688	0.697	0.688	0.689	0.697			
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Rent Inequality$_i$ ** Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93	-2.93

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is (log+1) of number of mentions of borough i in pages of job advertisements containing the word "*wanted* minus the (lo+1) of number of mentions of borough i in newspaper articles in month t using national or local newspapers available on the *British Newspaper Archive*. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E.5 Increase in State Capacity

In this Appendix Section we display several additional results and robustness checks for our analysis of channels presented in Section 6 of the main text.

We start with presenting the findings of two investigations with respect to proxies for state capacity. As first proxy for state capacity, one can use public expenditure in boroughs for years 1868 and 1869 [pre-post elections], using data from Knatchbull-Hugessen (1869) and Knatchbull-Hugessen (1870). This data is only available at the yearly level, which considerably restricts the statistical analysis. The results are displayed in Table A52 below. The estimates suggest no effect of the reform on this proxy of state capacity.

Further, the following Table A53 performs a similar analysis as in Table A52 but focuses on public deficits instead of public spending as proxy for increased state capacity building. Similarly to the conclusion on public expenditures, we do not detect any effect of enfranchisement. These findings are in line with the discussion in the main text, highlighting several key caveats for which our research design/data is unlikely to detect any state capacity effects in the short-run.

Table A52: Channels: Democracy and State Capacity - Public Expenditure

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: Public Expenditure$_{it}$	OLS	OLS	RF	RF	2SLS	2SLS
(log) Electorate$_{it}$	-0.00320	-0.00749			-0.0657	-0.0988
	(0.0712)	(0.0746)			(0.0797)	(0.104)
(log) Eligible Householders$_{it}$			-0.0548	-0.0871		
			(0.0663)	(0.0905)		
Observations	272	272	272	272	272	272
R-squared	0.991	0.991	0.991	0.991	-	-
1st stage F-Stat	-	-	-	-	130	131
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes
Rent Inequality$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes

Notes: The unit of observation is borough i in year t (where t corresponds to years 1868 and 1869). The full sample covers 136 boroughs. OLS (2SLS) estimates are reported in columns 1-4 [5-6]. The dependent variable is the (log) public expenditure in the borough in year t (Knatchbull-Hugessen (1870)). The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A53: Channels - Democracy and State Capacity - Deficit

Dep. Variable: Public Deficit$_{it}$	(1) OLS	(2) OLS	(3) RF	(4) RF	(5) 2SLS	(6) 2SLS
(log) Electorate$_{it}$	-3,280	-849.3			-6,541	-1,979
	(2,870)	(2,463)			(5,702)	(3,257)
(log) Eligible Householders$_{it}$			-5,452	-1,744		
			(4,744)	(2,914)		
Observations	272	272	272	272	272	272
R-squared	0.776	0.780	0.778	0.781	-	-
1st stage F-Stat	-	-	-	-	130	131
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Average Rents$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes
Rent Inequality$_i$ ** Time FEs*	No	Yes	No	Yes	No	Yes

Notes: The unit of observation is borough i in year t (where t corresponds to years 1868 and 1869). The full sample covers 136 boroughs. OLS (2SLS) estimates are reported in columns 1-4 [5-6]. The dependent variable is the public deficit in the borough i in year t (Knatchbull-Hugessen (1870)), computed as difference between expenditure and revenues in the borough i. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E.6 Heterogeneous Effects – Market Potential

In this and the following Appendix sections, we provide further information and display the tables of the heterogeneous effects analysis, summarized in the main text in Section 6.4. We start with the analysis of market potential as a magnifying factor of the economic dividend of the Second Reform Act. For each borough, we compute a simple market-potential index (Harris, 1954) which captures how close a given borough is to large markets with many potential consumers:

$$Market\ Potential_i = \sum_{\forall j \neq i} \frac{Population_j}{Distance_{ij}} \tag{4}$$

where $Population_j$ indicates the population of borough j in year 1866, $Distance_{ij}$ represent the distance between borough i and borough j.

Below we explore whether the effects of enfranchisement are magnified by greater market opportunities of a borough. Corresponding estimates are displayed in Tables A54 and A55. In line with the evidence presented in Section 6.2, we conclude that both the overall pacifying effect as well as the pro-growth impact of the Second Reform Act are magnified for towns in regions with larger marker potential.

Table A54: Democracy and Social Violence - Heterogeneous Effect - Market Potential

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) RF	(4) RF
(log) Electorate$_{it}$	-0.131***	-0.119***		
	(0.0362)	(0.0382)		
(log) Electorate$_{it}$ * *(log) Market Potential*$_i$	-0.131***	-0.240***		
	(0.0388)	(0.0838)		
(log) Eligible Householders$_{it}$			-0.125***	-0.133***
			(0.0424)	(0.0491)
(log) Eligible Householders$_{it}$ * *(log) Market Potential*$_i$			-0.146***	-0.267***
			(0.0443)	(0.0916)
Observations	4,416	4,416	4,416	4,416
R-squared	0.392	0.412	0.391	0.412
Borough FEs	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	No	Yes
Rent Inequality$_i$ * *Time FEs*	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. The variable *Market Potential*$_i$ is computed following the approach described in Section E.6. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A55: Democracy and Economic Growth - Heterogeneous Effect - Market Potential

Dep. Variable: (log) City Mentions in Newsp. Ads$_{it}$	(1) OLS	(2) OLS	(3) RF	(4) RF
(log) Electorate$_{it}$	0.166***	0.171***		
	(0.0431)	(0.0444)		
(log) Electorate$_{it}$ * *(log) Market Potential*$_i$	0.144***	0.0997*		
	(0.0526)	(0.0521)		
(log) Eligible Householders$_{it}$			0.191***	0.191***
			(0.0471)	(0.0518)
(log) Eligible Householders$_{it}$ * *(log) Market Potential*$_i$			0.167***	0.126**
			(0.0626)	(0.0608)
Observations	4,416	4,416	4,416	4,416
R-squared	0.952	0.954	0.953	0.954
Borough FEs	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes
Average Rents$_i$ * *Time FEs*	No	Yes	No	Yes
Rent Inequality$_i$ * *Time FEs*	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. OLS (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is (log+1) of number of mentions of borough i in pages of job advertisements in month t using national or local newspapers available on the *British Newspaper Archive*. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. The variable *Market Potential*$_i$ is computed following the approach described in Section E.6. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E.7 Heterogeneous Effects – Ethnic Polarization

Another key dimension of potential heterogeneity is the local population composition: during the 1860s, social tensions were substantial between the Anglican Englishmen and Catholic immigrant laborers from Ireland. The so-called "Murphy riots" (see Arnstein, 1975) were a manifestation of these tensions. Hence, we expect that the pacifying effect of the enfranchisement of part of the working class may be larger in areas with greater social tensions. In previous work it has been found that ethnically polarized societies are on average more likely to be plagued by inter-group conflict (see Montalvo and Reynal-Querol, 2005 and Esteban, Mayoral and Ray, 2012). To investigate whether ethnic polarization indeed magnifies our enfranchisement effects, we compute, for each borough, the level of ethnic polarization, following the approach of Montalvo and Reynal-Querol (2005), and distinguishing between two groups: 1) Individuals born in England or Wales; 2) Irish born individuals. As shown in Table A56 below, in areas with higher ethnic polarization political reform has –as expected– a greater pacifying potential.

Table A56: Democracy and Social Violence - Heterogeneous Effect - Polarization Irish vs English/Welsh

Dep. Variable: Social Violence$_{it}$	(1) OLS	(2) OLS	(3) RF	(4) RF
(log) Electorate$_{it}$	-0.114***	-0.110**		
	(0.0390)	(0.0441)		
(log) Electorate * Polar Irish vs UK-WAL	-0.814*	-0.886*		
	(0.438)	(0.485)		
(log) Eligible Householders			-0.106**	-0.110**
			(0.0446)	(0.0520)
(log) Eligible Householders * Polar Irish vs UK-WAL			-1.173**	-1.348***
			(0.459)	(0.491)
Observations	4,416	4,416	4,416	4,416
R-squared	0.390	0.398	0.390	0.398
City FEs	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes
Average Gross Rental * Month FEs	No	No	No	No
Average Gross Rental * Month FEs * Market Potential	No	No	No	No
Gini Gross Rental * Month FEs	No	Yes	No	Yes
Gini Gross Rental * Month FEs * Market Potential	No	Yes	No	Yes

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate*$_{it}$ corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders*$_{it}$ corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. The variable *Polarization*$_i$ indicates the polarization index in borough i computed using two groups: Irish-born inhabitants and England and Wales-born inhabitants (data from the 1861 Census (IPUMS (2020))). Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E.8 Heterogeneous Effects – Types of Social Violence

In this Appendix Section we investigate what types of social violence are affected by the enfranchisement of the Second Reform Act. The historical literature on the "Age of Reforms" in Victorian England (see Section 2) highlights that some of the social tensions during the 1860s were linked to claims for political inclusion (think e.g. of some of the "Reform League" demonstrations), while some riots were linked to ethno-religious competition and social conflict (think e.g. of the "Murphy Riots"). One may expect franchise extension to reduce (almost mechanically) the unrest linked to claims for participation and enfranchisement (as many requests have been satisfied by the Second Reform Act). At the same time, the enfranchisement gave a political voice to the urban working class, a part of it being from Irish origin, which similarly could have reduced social tensions by fostering political inclusion of all major ethnic groups (see e.g. Mueller and Rohner, 2018). Hence, we expect both a reduction in *political* and *ethno-religious* types of social conflict.

Further, to the extent that enfranchisement boosted the economy (see Section 6.2), one may expect a higher opportunity cost of engaging in social unrest in booming boroughs. This could attenuate the risk of all kinds of social conflict, not only the aforementioned political and ethno-religious ones, but also others, such as *economic* types of social conflict.

While it is difficult to pin-down the exact type of social conflict for a given incident, in what follows we carry out an explorative, keyword-based exercise going in this direction. In particular, the presence of the following keywords (nearby the text string related to social conflict) classifies a given incident into the three aforementioned categories of social conflict types:

- Political: representation, suffrage, franchise, voters, democracy
- Religious: protestant, church, papal, chapel, priest, catholic, clergy, murphy, religion, religious
- Economic: strike, workers, salaries, unemployed, workmen, labourers, unemployed, ironworkers, workpeople, labour, trade unions

As a next step we run the baseline specification of the benchmark Table 1, but with as dependent variable the three measures of specific types of social violence. The results are displayed in Tables A57, A58 and A59, respectively. We find that both ethno-religious and economic types of social violence tend to be reduced by enfranchisement.

Moreover, we apply an alternative method to classify the social conflict episodes into different categories. The method used is based on the “BART-large MNLI” model developed by Facebook (and available on the HuggingFace platform). This model uses a Bidirectional and Auto-Regressive Transformer (BART) algorithm (Lewis, Liu, Goyal, Ghazvininejad, Mohamed, Levy, Stoyanov and Zettlemoyer, 2019) pre-trained on a large Multi-Genre Natural Language Inference (MNLI) dataset containing a collection of 433.000 sentence pairs annotated with

Table A57: Democracy and Social Violence - Politics-related events - Bag-of-words model

Dep. Variable: Social Violence [Political-BagsOfWords]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.00703	-0.0125	-0.0125				0.000409	-0.00942	-0.00941
	(0.00875)	(0.00877)	(0.00895)				(0.0118)	(0.0122)	(0.0121)
(log) Eligible Householders$_{it}$				0.000355	-0.00844	-0.00847			
				(0.0102)	(0.0109)	(0.0109)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.088	0.100	0.104	0.088	0.099	0.104	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.01	.01	.01	.01	.01	.01	.01	.01	.01

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a politics-related violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Sections 3.1 and E.8. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

textual entailment information. This method has become increasingly popular, as it is capable of allocating text-based information into an unseen set of labels (i.e., categories). In other words, the "BART-large MNLI" model has never been specifically trained to classify social conflict episodes, but it has been designed to produce a zero-shot label (i.e., without pre-training on class-labelled training data) of any text-based variable by exploiting Natural Language Inference (NLI) models proposed by Yin, Hay and Roth (2019).

In our case, we have defined 3 different labels (economy, politics and religion) to be recognized in parallel, by setting up the classifier as multi-label. With respect to the traditional bag-of-words method (i.e. distinguishing different conflict types using keywords specified by the researcher), the "BART-large MNLI" model offers two main advantages. First, it reduces the potential leeway associated with the selection of keywords in the bag-of-words approach. Secondly, it allows an exclusive categorization of events into one single category. Put differently, the output of the model corresponds to the probabilities that a given input falls into each category. It is hence possible to assign the social conflict-string to the category that has received the highest probability.

The results obtained with this novel, alternative method of distinguishing types of social conflict events are presented in three new Tables focusing, respectively, on politics-related (see

Table A58: Democracy and Social Violence - Ethno-Religious-related events - Bag-of-words model

Dep. Variable: Social Violence [Religious-BagsOfWords]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.114***	-0.109***	-0.112***				-0.145***	-0.144***	-0.142***
	(0.0332)	(0.0345)	(0.0358)				(0.0460)	(0.0503)	(0.0493)
(log) Eligible Householders$_{it}$				-0.126***	-0.129***	-0.128***			
				(0.0399)	(0.0450)	(0.0443)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.354	0.362	0.366	0.354	0.362	0.365	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.072	.072	.072	.072	.072	.072	.072	.072	.072

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a religious-related violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Sections 3.1 and E.8. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A60), religious-related (see Table A61) and economic-related (see Table A62) social conflict episodes. Overall, our results are consistent with the notion that the franchise extension has resulted in a broader pacifying effect across several types of social conflict events. Integrating economically and electorally the nascent Irish Catholic middle class proved of key importance to build a peaceful and stable political system for the years to come.

Table A59: Democracy and Social Violence - Economic-related events - Bag-of-words model

Dep. Variable: Social Violence [Economic-BagsOfWords]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.0334	-0.0504**	-0.0516**				-0.0313	-0.0679**	-0.0671**
	(0.0215)	(0.0202)	(0.0209)				(0.0320)	(0.0321)	(0.0313)
(log) Eligible Householders$_{it}$				-0.0271	-0.0609**	-0.0604**			
				(0.0277)	(0.0283)	(0.0278)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.174	0.193	0.197	0.173	0.193	0.197	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.027	.027	.027	.027	.027	.027	.027	.027	.027

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a economic-related violent event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Sections 3.1 and E.8. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A60: Democracy and Social Conflict - Politics-related events - Zero-Shot Text Classification

Dep. Variable: Social Violence [Political-ZeroShot]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.0987***	-0.100***	-0.103***				-0.133***	-0.146***	-0.145***
	(0.0306)	(0.0330)	(0.0334)				(0.0408)	(0.0460)	(0.0450)
(log) Eligible Householders$_{it}$				-0.116***	-0.131***	-0.130***			
				(0.0352)	(0.0404)	(0.0398)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.351	0.354	0.360	0.352	0.354	0.360	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.112	.112	.112	.112	.112	.112	.112	.112	.112

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a politics-related violent (classified using the zero-shoot algorithm described in the text) event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A61: Democracy and Social Conflict - Religious-related events - Zero-Shot Text Classification

Dep. Variable: Social Violence [Religious-ZeroShot]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.0982***	-0.100***	-0.102***				-0.119***	-0.129***	-0.128***
	(0.0285)	(0.0298)	(0.0309)				(0.0400)	(0.0439)	(0.0432)
(log) Eligible Householders$_{it}$				-0.103***	-0.116***	-0.115***			
				(0.0348)	(0.0391)	(0.0388)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.324	0.330	0.334	0.323	0.330	0.333	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.081	.081	.081	.081	.081	.081	.081	.081	.081

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a religious-related violent (classified using the zero-shoot algorithm described in the text) event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A62: Democracy and Social Conflict - Economics-related events - Zero-Shot Text Classification

Dep. Variable: Social Violence [Economic-ZeroShot]$_{it}$	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF	(7) 2SLS	(8) 2SLS	(9) 2SLS
(log) Electorate$_{it}$	-0.0858**	-0.0846**	-0.0835**				-0.0869*	-0.0846*	-0.0853*
	(0.0340)	(0.0353)	(0.0360)				(0.0445)	(0.0474)	(0.0471)
(log) Eligible Householders$_{it}$				-0.0753*	-0.0758*	-0.0768*			
				(0.0386)	(0.0422)	(0.0423)			
Observations	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416	4,416
R-squared	0.362	0.365	0.370	0.361	0.364	0.369	-	-	-
1st stage F-Stat	-	-	-	-	-	-	232	194	184
Borough FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Average Rents$_i$ * Time FEs*	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
*Rent Inequality$_i$ * Time FEs*	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.173	.173	.173	.173	.173	.173	.173	.173	.173

Notes: The unit of observation is borough i and month t. The sample covers 184 boroughs over the period 1868-1869. LPM (2SLS) estimates are reported in columns 1-6 [7-9]. The dependent variable is a dummy that takes a value of 1 if a economy-related violent (classified using the zero-shoot algorithm described in the text) event was observed in borough i and month t. The social violence data was constructed using national or local newspapers available on the *British Newspaper Archive*, following the approach described in Section 3.1. The variable *(log) Electorate$_{it}$* corresponds to the electorate pre and after the Elections of 1868 in a borough i. The variable *(log) Eligible Householders$_{it}$* corresponds to the number of householders with rental value above £10 for the period before the Elections of 1868 and the number of householders living in a house with rental value above £10 and between £4 and £10 (who were previously banned from voting but enfranchised in the Second Reform Act) for the period after the Elections of 1868. Borough-level rent-based variables have been computed using data from House of Commons (1866*a*). Robust standard errors clustered at the borough level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F. Appendix: Supplementary data – Lists of newspapers used

The current Appendix Section is devoted to supplementary data. In particular, Table A63 displays all 471 national or local newspapers used to construct our novel social conflict data. The total number of articles containing at least one social conflict-related keyword available during our sample period is reported in parenthesis.

Table A63: Lists of newspapers used

London Evening Standard (# 5984)	Aberdeen Journal and North Advertiser (# 794)	Fallkirk Herald (Agus Dunedin) (# 594)	Alfreton Journal (# 367)	Grantham Journal (# 883)	Northern Herald (# 598)	Reading and East Berks Chronicle (# 97)
Sun (London) (# 3964)	Uxbridge & W. Drayton Gazette (# 767)	Worcester Journal (# 301)	Inverness Courier (# 588)	Hills and Gloucestershire Standard (# 188)	Cheltenham Mercury (# 198)	Public ledger and Daily Advertiser (# 93)
The Guardian (# 3644)	Norwich Mercury (# 644)	Norfolk Chronicle (# 301)	Hampshire Chronicle (# 598)	Morning Telegraph, and General Advertiser for the... (# 361)	Galway Vindicator, and Connaught Advertiser (# 307)	Farmer's Gazette and Journal for Practical Farmers... (# 86)
Morning Post (# 5495)	Weekly Pennsylvania Journal (# 681)	Derby Mercury (# 598)	Oxford University and Borough Herald (# 364)	The Portsmouth Times and Sun Ruffalo Advertiser (# 186)	Gateshead and Northumberland Advertiser, and Newc... (# 307)	Grain's Liverpool General Advertiser (# 87)
Liverpool Mercury (# 1438)	Manchester Times (# 677)	Oxford Times (# 495)	Bristol Mercury and Gloucestershire Advertiser (# 370)	Watford Observer (# 186)	Worcester-shire Mayor Gazette, and General Advertiser... (# 393)	Waterford Mirror and Munster Visitor (# 67)
Leeds Mercury (# 1464)	Stoke (# 972)	Bobby Herald and Northumbria Advertiser (# 446)	Evening and Mail Newspaper (# 377)	Carlow Post (# 279)	Berkshire Advertiser and North Wilts Chronicle (# 187)	Eastbourne Gazette (# 86)
Liverpool Daily Post (# 3611)	Isleworth Chronicle (# 689)	Cardiffe Patriot (# 495)	Essex Borough Post (# 376)	The Examiner (# 376)	Bentley Gazette (# 187)	Penrose Worchester and Lodge Gazette (# 85)
Yorkshire Post and Leeds Intelligencer (# 3286)	Bristol Mercury (# 668)	Gloucestershire Chronicle (# 482)	The Salisbury Times (# 375)	Gosey Express, Bromley Hill, Beckenham, Kidbo... (# 276)	Gateshead Request, and Ware's Wheelbarrows Advert... (# 186)	Middlesex County Times (# 84)
Sheffield Independent (# 3276)	Halifax Courier (# 646)	Shire Gazette (# 466)	Portsmouth Gazette (# 375)	General Bonnard Gazette (# 277)	Dublin Medical Press (# 186)	Hastings and St Leonards Observer (# 84)
Edinburgh Evening Courant (# 3259)	Gravesend Telegraphand Clyde Shipping Gazette (# 664)	Gravesend Advertiser (# 486)	Royal Cornwall Gazette (# 375)	Thame Chronicle (# 376)	Berkshire Advertiser (# 185)	Oxfordshire Weekly News (# 83)
Sheffield Daily Telegraph (# 2867)	Coventry Standard (# 646)	Leicester Chronicle (# 466)	Portsmouth Times and Naval Gazette (# 376)	Wrexham Journal (# 236)	Bentley Guardian (# 184)	Bedford Journal (# 82)
Morning Advertiser (# 2836)	Newry Examiner and South Advertiser (# 661)	Exonerated Bonnard Journal (# 461)	Tipperary Vindicator (# 375)	Hampshire Advertiser (# 372)	Berkshire Telegraph and General Advertiser for th... (# 184)	Buxton Advertiser (# 78)
Bristol Times and Mirror (# 2371)	Newcastle (# 657)	Times of India (# 459)	Durham Chronicle (# 371)	Aberdeen Pantheon (# 372)	Sporting Life (# 183)	Abbeyfeasane and Rhuddlan Advertiser (# 77)
Civil Examiner (# 2141)	Cardiff and Merthyr Guardian, Glamorgan, Monmouth... (# 633)	Bath Chronicle and Weekly Gazette (# 477)	Croydon Chronicle and East Surrey Advertiser (# 370)	North London Chronicle (# 371)	Dartmouth Chronicle (# 183)	New Sussex Gazette (# 73)
Dublin Evening Mail (# 2115)	Essex Standard (# 625)	Derbyshire Advertiser and Journal (# 476)	Bridgnorth Journal and South Shropshire Advertiser (# 364)	Leamington Advertiser, and Bath's Star of Victor... (# 371)	Essex Telegraph and Gloucestern General Advertise... (# 179)	Isle of Wight Observer (# 73)
London Daily News (# 2087)	Reliant Herald (# 622)	North British Agriculturist (# 471)	Leicester Journal (# 367)	Roscommon Messenger (# 176)	Bromley Herald and General Advertiser for Bush... (# 178)	Cheltenham Journal and Gloucestershire Nationalis... (# 73)
Pontefract Mail (# 2073)	Yorkshireman (# 620)	Kendal Gazette (# 454)	Kings Lynn Chronicle (# 367)	Wakefield Express and General Advertiser (# 364)	Basingstoke Advertiser and Farm Press (# 178)	Coventry Guardian (# 73)
Globe (# 1960)	Lloyd's Weekly Newspaper (# 616)	Wellington Journal (# 471)	Bridlemire Observer (# 367)	Reigate Patriot (# 266)	Anglesea Bonnard Observer and Lincoln Gazette (# 177)	Chronicle News (# 69)
Southern Reporter and Civil Commercial Courier (# 1860)	Leeds Times (# 611)	Hull Packet (# 471)	Anglesea Guardian (# 367)	Horsham, Petworth, Midhurst and Steyning Express (# 362)	Kilgrim Advertiser (# 177)	Surrey Advertiser (# 68)
Civil Commanders (# 1793)	Dublin Daily News (# 601)	Salisbury and Winchester Journal (# 471)	Chelsea Free and General Advertiser (# 346)	Cornish Advertiser (# 358)	Aberystwyth Times (# 177)	Ruin Police Reporter (# 86)
Manchester Courier and Lancashire General Advertis... (# 1780)	Reynolds's Newspaper (# 601)	Reading Mercury (# 446)	Cheltenham Times and Gloucestershire Herald (# 346)	Battersea Telegraph (# 357)	Kilkenny Moderator (# 176)	Era Agrarian (# 68)
Glasgow Herald (# 1745)	Lancaster Gazette (# 600)	Bamsted Mercury (# 446)	York Advertiser (# 343)	West London Observer (# 336)	Wexford People (# 176)	Belfast Mercantile Register and Weekly Advertiser (# 58)
Birmingham Daily Post (# 1736)	Ann's Reminders Gazette (# 593)	Bury Free Press (# 418)	Essex Herald (# 341)	London Evening Post (# 353)	Cumberland & Westmoreland Herald (# 176)	Essex Weekly Times (# 57)
Bradford Observer (# 1707)	Durham County Advertiser (# 586)	Bonaparte's Recruiter (# 446)	Falmouth (# 341)	Friend of India and Statesman (# 351)	Otley News and Wharfedale Advertiser (# 173)	Exeter's Eleven Times (# 56)
Nottingham Journal (# 1700)	Londonderry Standard (# 584)	Staffordshire General and Commercial Gazette, Ad... (# 464)	Liverpool London Globe (# 339)	Thame Advertiser (# 351)	Newcastle Guardian and Tyne Mercury (# 171)	St. Andrews Gazette and Fifeshire News (# 55)
Pall Mall Gazette (# 1651)	Rochdale Observer (# 585)	Wrexham Advertiser (# 461)	Kerry Evening Post (# 351)	Southern Reporter (# 346)	Bathroom & West Gyrfery Bugle, or, South Wales... (# 167)	Wrexham, Guardian and Denbighshire and Flintshire A... (# 54)
Newcastle Journal (# 1631)	Congleton & Macclesfield Mercury, and Cheshire Ga... (# 580)	Margaret Advertiser (# 458)	Liverpool Times (# 351)	Hertford Mercury and Reformer (# 346)	Oxfordshire Express and West Essex Journal (# 167)	Ulster Gazette (Advertising, Herald of Business and... (# 51)
York British Daily Mail (# 1577)	Reverend Advertiser (# 558)	Derry Journal (# 436)	Northern Standard (# 346)	Bedfordshire Advertiser (# 346)	Bolton Observer and East Manchesterlive Advertiser (# 166)	Bolly Avenue Herald (# 166)	North Riders (# 51)
Saunders's News Letter (# 1556)	Hampshire Telegraph (# 553)	Chelmsford Chronicle (# 437)	Ryecroft News (# 345)	Bridgwater Journal, and General Advertiser for the... (# 365)	Colindale Chronicle (# 164)	Bothwell Times (# 49)
Birmingham Daily Gazette (# 1531)	Clerkenwell News (# 552)	Cheshire Observer (# 437)	The Tewkesbury Register, and Agricultural Gazette... (# 346)	Kilkenny Journal (# 344)	Derry and Navy Gazette (# 358)	Merrylle Express (# 44)
The Evening Express (# 1530)	Bonaparte Herald (# 536)	Western Mercury and County Advertiser (# 455)	Thames County Times (# 360)	Advertiser Journal (# 343)	The Star (# 336)	Bellbrook Advertiser (# 43)
Freeman's Journal (# 1514)	Abergavenny Advertiser (# 536)	Taunton Gazette, and Western Advertiser (# 434)	Taunton Gazette (# 345)	Croydon (# 342)	The Arbs (# 357)	Bengaluru Advertiser and Hub of the Waterfronts(... (# 43)
Dublin Evening Post (# 1513)	Northern Standard (# 533)	Lyle's Falmouth Packet and General Advertiser (# 434)	Waterford Chronicle (# 344)	East London Observer (# 336)	Wellington (# 357)	Cheltenham Looker-On (# 42)
Bristol Daily Post (# 1498)	Birmingham Journal (# 531)	Ross Gazette (# 413)	Bury Times (# 341)	Home-ward Mail from India, China and the East (# 340)	The Economy of the Poor, and County of Wexford Advoc... (# 356)	Glamorgan Gwent Chronicle and North Staffordshire Reporter,... (# 41)
Glasgow Evening Citizen (# 1386)	Newry Reporter (# 531)	Norfolk Chronicle (# 448)	Suffolk and Essex Free Press (# 337)	Barnard Castle News (# 339)	Barnsley Advertiser (# 356)	Irish Environmental Gazette (# 41)
Dublin Daily Express (# 1332)	Killarney Journal, and Lakener Commercial and Lis... (# 506)	Chapeise Weekly Advertiser (# 448)	Arbroath Guide (# 337)	Bethesy Advertiser (# 335)	Barnes Gazette (# 353)	Torquay Times, and South Devon Advertiser (# 38)
Belfast News-Letter (# 1305)	Chester Chronicle (# 506)	Gainsborough Independent Press (# 447)	Falkirk Herald (# 336)	Leigh Chronicle and Weekly Elector Advertiser (# 335)	Tadcaster Post, and General Advertiser for Grimm... (# 336)	Wear Banner Times (# 37)
Western Daily Press (# 1280)	Barnsley Chronicle, etc (# 351)	County Advertiser & Herald for Staffordshire and W... (# 446)	Basingstoke Express (# 336)	Brampton Weekly News (# 328)	Perry Signs and Notes Gazette, Northern Reviews a... (# 343)	Medway and Kingsland Gazette (# 37)
Bristol Observer (# 1230)	Reading Mercury (# 350)	Oxford Chronicle and Reading Gazette (# 441)	Blythe Life in London and Society Chronicle (# 334)	Greenwich and Woolwich Times (# 327)	Tadcaster Parish Gazette, and Mercury Reports (# 336)	The Dublin Packet (# 35)
Newcastle Daily Chronicle (# 1219)	Abergavenny Chronicle (# 504)	Chelmsford Times (# 440)	Morpeth Herald (# 336)	Otley Times, and Aegithlee Advertiser (# 326)	Tadcaster Advertiser and Wakefieldridge Newsletter (# 340)	Sublee Advertiser (# 33)
Shipping and Mercantile Gazette (# 1128)	Clare Journal, and County Advertiser (# 504)	North Devon Journal (# 441)	Frant Chronicle Gazette (# 330)	Reading Mercury Weekly News (# 324)	Tavistock Independent (# 336)	Surrey Observer (# 32)
The Nation Weekly Reporter, and Ballybridge audit... (# 1114)	Kendal Independent (# 501)	Bury and Norwich Post (# 436)	Renfrew Journal and Provost Advertiser (# 326)	Liverpool Mail (# 336)	Cumberland Herald (# 31)	Tamworth and Alrewas-on-Weedy Herald (# 30)
Bell's Weekly Messenger (# 1069)	Hampshire Advertiser (# 351)	Brighton Argus and Sussex Journal (# 414)	St. Ives Chronicle and Advertiser (# 333)	Neuse & Aldborough Gazette and Weekly Chronicle of t... (# 321)	Berkshire Constitutional & Journal (# 335)	Hexe Mars Times (# 30)
Irish Times (# 1063)	Pall (# 448)	Bradly's Observer Advertiser and General Intelig... (# 413)	Windsor and Eton Express (# 332)	Bexhill Mercury (# 324)	Yarm Herald (# 335)	Glasgow Free Press (# 27)
Primrose Herald (# 1015)	Exeter and Plymouth Gazette (# 446)	Renfrewshire Independent (# 413)	Worcestershire Chronicle (# 331)	Frome Times (# 331)	Wiltshire Independent (# 333)	County Cavan Chronicle (# 26)
Wigan Observer and District Advertiser (# 1014)	The Irishman (# 541)	Oxfordshire Telegraph (# 436)	Kilgo Champion (# 310)	Glossop Record (# 331)	Herefordshire Express and General Advertiser (# 330)	Buxton Herald (# 26)
Dundee Advertiser (# 1013)	Cardiff Times (# 542)	West Somerset Free Press (# 436)	Cumberland & Dunlop Herald (# 331)	Gainsborough Weekly Reporter (# 330)	Renfrewshire Gazette (# 330)	Reading Times (# 26)
Belfast Morning News (# 1010)	Armagh Guardian (# 540)	Warrenford Mail (# 436)	Folkestone Express, Sandgate, Elsmere Park Hydro... (# 327)	Croydon's Weekly Standard (# 330)	Meath Herald and Cavan Advertiser (# 328)	Wrexham News (# 25)
Daily Telegraph & Courier (London) (# 976)	Rochdale Pilot Chronicle (# 527)	Cambridge Chronicle and Journal (# 428)	Shipton Mallet Journal (# 330)	Witney Herald, and Weekly Advertiser and Gazette... (# 327)	Wells Journal (# 326)	Worcestershire Herald (# 25)
Morning and Daily Weekly Mail (# 975)	Cheltenham Examiner (# 516)	Gravesend Reporter, North Kent and South Essex Ad... (# 427)	Wakefield Standard (# 314)	Exeter and Plymouth Gazette Daily Telegraph (# 327)	Witney Express and Oxfordshire and Midland Country... (# 325)	Kirby's List (# 25)
Banner of Ulster (# 974)	Bolton Evening News (# 515)	Northern Daily Leaders (# 426)	The Era (# 313)	Warwickshire Times and Home Rule Herald (# 327)	Salisbury Reporter (# 324)	Wentworth Gazette (# 25)
The Nation Weekly Reporter, and Ballybridge audit... (# 958)	Surrey Comet (# 506)	The Ipswich Journal (# 410)	Norwich and Warwickshire Advertiser (# 306)	Canterbury Journal, Kentish Times and Farmers' Ga... (# 320)	Yeovily Express (# 319)	Barnet or Amwerso Comes (# 25)
Primrose Herald Chronicle (# 945)	Gloucester Times and Cotswold Advertiser (# 505)	Scarborough Gazette (# 410)	Cromarty Herald (# 306)	Yorkshire Times and Strathspey Advertiser (# 320)	Hereford Times (# 318)	The Graphic (# 24)
Ulster Examiner and Northern Star (# 939)	Hull and Eastern Counties Herald (# 502)	Worcestershire (# 410)	Yeverley Weekly News and General Advertiser (# 304)	Hope Constitution (# 318)	Teesdale Mercury (# 317)	Tattersalls Journal and Directory (# 7)
Essex Telegraph (# 936)	Hull & County Weekly News, and York Chronicle, (# 502)	Ellesmere Times Guardian (# 409)	Wolverhampton Chronicle and Staffordshire Advertis... (# 306)	Bexley Gazette (# 315)	Bournemouth & Dorset Gazette (# 315)	Ossett Times (# 7)
Bradford Daily Telegraph (# 934)	Montrose, Arbroath and Brechin review, and Bothe... (# 526)	Montrose Standard (# 490)	Bridport News (# 301)	Bude Herald and Boats Gazette (# 310)	Bewdley Reporter (# 313)	Courier of the Exchange (# 7)
Lincolnshire Chronicle (# 876)	Aberdeen Free and Journal (# 523)	The Suffolk Chronicle, or Weekly General Advertis... (# 417)	Tipton Constitution (# 301)	Watford Independent (# 313)	Kidderminster Times and Advertiser for Bewdley &S... (# 311)	Liverpool Courier and Commercial Advertiser (# 5)
Leicester Mail (# 870)	Berkshire Chronicle (# 515)	Bedfordshire Mercury (# 417)	Surrey Courier (# 300)	Limerick Reporter (# 314)	Rugby Advertiser (# 310)	Southern Times and Dorset County Herald (# 3)
Western Times (# 867)	Bristol Journal (# 507)	Tipperary Free Press (# 410)	Jedburgh and Borders Herald (# 297)	Ashbourne Telegraph (# 314)	Bridgwater Journal (# 309)	Middlesex Chronicle (# 25)
Bedfordshire Times and Independent (# 860)	Leicester Guardian (# 536)	South London Press (# 410)	Tower Hamlets Independent and East End Local Paper... (# 296)	Derry Protestant Paper (# 307)	Yorkshire Mercury Gazette (# 308)	Palmitec Advertiser (# 3)
Brighton Herald and General Advertiser for the Co... (# 859)	Preston Chronicle (# 536)	North Wales Chronicle (# 41)	Barnes Chronicle (# 296)	Bolton Chronicle (# 306)	The Cornish Telegraph (# 308)	Chatham News (# 3)
Express (London) (# 846)	Tiverton Gazette (# 513)	Ulverston Mirror and Furness Reflector (# 407)	Bude Herald (# 293)	Horsehorn News (# 306)	The Worcester express, or, weekly commercial & agri... (# 307)	Barnstaple & South Devon chronicle (# 2)
Staffordshire Advertiser (# 844)	East Kent Gazette (# 513)	Lowlandberry Journal (# 486)	Essex Express (# 296)	Halifax Advertiser (# 296)	Arundl Mercury (# 306)	Graphic (# 1)
Dublin Daily Gazette (# 842)	Gloucester Journal (# 498)	West Middlesex Herald (# 403)	Bristol Mirror (# 293)	Brighton Gazette (# 295)	Barnstaple Journal (# 306)	
Dublin Weekly Nation (# 843)	Newcastle Courant (# 498)	Elgin Courant, and Morayshire Advertiser (# 400)	Penrith Observer (# 288)	Lewes Times and Advertiser (# 303)	Bridlington Free Press (# 306)	
Nottinghamshire Guardian (# 833)	The Sportsman (# 364)	Mercer Chronicle (# 399)	Frant & Rye Chronicle and Eros Bulwr (# 288)	Nuneaton Advertiser (# 303)	Wakefield Constitution (# 301)	
Western Morning News (# 827)	Bexley Chronicle and Reports Journal (# 596)	Northampton Mercury (# 397)	Commercial Journal (# 286)	Ribchchester (# 303)	Northcoates News (# 305)	
Bolton Chronicle (# 813)	Express and Echo (# 596)	Oxford Journal (# 397)	Longford Journal (# 286)	Dover Packet Post (# 301)	Maidstone Mercury (# 305)	

G. Appendix: Supplementary data – Additional rents statistics

The current Appendix section is devoted to supplementary data on rents, listed for each borough separately. In particular, Table A64 reports average rents, rents inequality (measured using the Gini Index), as well as the number of householders in total and per relevant rent bracket for each borough.

Table A64: Rents statistics by borough

Name	Ave. Rents	Gini Index	Num. Householders	HH <£50	HH £50-£100	HH >£100	Name	Ave. Rents	Gini Index	Num. Householders	HH <£50	HH £50-£100	HH >£100	Name	Ave. Rents	Gini Index	Num. Householders	HH <£50	HH £50-£100	HH >£100	Name	Ave. Rents	Gini Index	Num. Householders	HH <£50	HH £50-£100	HH >£100
Abingdon	17.55	0.46	970	609	310	361	Derby	19.66	0.48	7,869	5,250	4,955	2,619	Ludlow	20.41	0.45	937	514	408	423	Scarborough	25.92	0.53	4,210	2,146	1,604	1,965
Andover	28.98	0.59	935	663	469	272	Devizes	25.63	0.52	916	542	471	374	Lymington	26.29	0.53	870	462	366	408	Shaftesbury	31.70	0.64	1,362	1,130	437	452
Aldeburgh-upon-Sea	20.78	0.41	5,062	4,204	3,406	878	Dorchester	30.66	0.59	4,525	406	377	3,919	Macclesfield	13.97	0.37	6,541	5,281	3,635	1,055	Sheffield	18.25	0.45	38,494	28,558	26,047	10,160
Aldeburgh	25.73	0.6	5,260	3,060	1,401	1,300	Dorchester	24.70	0.51	828	381	363	447	Maidstone	23.48	0.47	4,153	1,960	1,839	2,193	Shrewsbury	36.63	0.58	5,106	3,197	2,516	1,959
Banbury	25.65	0.59	1,937	1,148	918	779	Dover	32.88	0.46	3,328	998	961	2,340	Maldon	21.02	0.5	1,206	886	769	320	Shrewsbury	23.55	0.48	3,301	2,246	2,080	1,067
Barnstaple	17.53	0.52	1,867	1,198	962	669	Droitwich	49.28	0.6	1,293	902	774	391	Malmesbury	34.52	0.61	1,187	835	462	352	South Shields	14.27	0.38	7,756	6,468	4,070	1,271
Bath	47.90	0.47	5,856	1,588	1,312	4,268	Durham	18.57	0.51	2,177	1,249	849	928	Malton	22.64	0.55	1,318	821	313	497	Southampton	30.07	0.39	7,276	512	738	6,764
Battersea	19.94	0.48	1,535	1,121	351	414	Evesham	23.48	0.51	833	504	428	329	Manchester	26.29	0.49	63,351	35,875	33,256	27,096	Southwark	36.88	0.44	24,662	7,051	6,736	17,649
Bedford	23.53	0.49	2,462	1,204	1,060	1,198	Exeter	30.82	0.52	5,789	3,023	2,462	2,766	Marlborough	26.80	0.51	713	442	361	271	St Ives	13.39	0.47	1,372	1,147	958	385
Berwick-upon-Tweed	19.91	0.52	2,402	1,894	352	508	Eye	33.35	0.58	1,287	980	453	307	Marlow, Great	31.39	0.6	1,040	681	407	359	Stafford	14.38	0.43	2,452	1,649	1,569	803
Beverley	24.90	0.52	2,015	1,377	1,102	638	Finsbury	48.49	0.32	41,107	1,520	1,405	39,908	Marylebone	86.78	0.29	35,961	456	413	35,138	Stamford	23.23	0.53	1,215	769	572	467
Bewdley	16.20	0.52	1,632	1,091	594	546	Flint	17.50	0.35	3,402	2,651	1,175	751	Merthyr Tydfil	25.55	0.59	14,911	12,963	9,910	1,948	Stockport	17.95	0.45	8,912	7,237	5,445	1,405
Birmingham	20.68	0.48	97,455	39,561	37,054	19,362	Frome	12.67	0.31	1,761	1,356	663	405	Melksham	28.80	0.62	1,167	638	454	529	Stockton-on-Tees	13.81	0.43	18,590	15,173	13,662	3,417
Blackburn	16.66	0.41	10,389	8,791	7,463	1,598	Gateshead	16.50	0.48	7,210	5,744	4,001	1,466	Monmouth	34.17	0.44	4,089	1,375	1,149	2,714	Stroud	20.60	0.56	6,927	5,446	2,421	1,483
Bodmin	28.56	0.57	1,006	620	385	377	Gloucester	25.80	0.49	2,961	1,203	1,152	1,606	Montgomery	19.60	0.54	3,453	2,459	1,980	984	Sunderland	18.74	0.47	15,108	10,758	8,807	4,427
Bolton	18.54	0.46	12,213	9,476	8,218	2,737	Grantham	25.52	0.48	2,008	1,308	1,154	700	Morpeth	32.03	0.58	1,585	1,110	564	455	Swansea	18.10	0.49	10,776	7,458	5,227	3,318
Boston	23.75	0.53	3,014	2,140	1,706	874	Great Grimsby	17.19	0.46	3,491	2,738	2,049	753	Newark	21.56	0.53	2,158	1,305	1,128	853	Tamworth	26.05	0.52	1,866	1,137	918	569
Bradford	19.85	0.51	22,796	16,626	13,617	6,170	Greenock	27.14	0.42	10,376	5,160	4,986	15,216	Newcastle-under-Lyme	14.05	0.46	2,455	1,826	1,549	629	Taunton	19.91	0.52	2,754	1,611	1,372	1,123
Bridgnorth	20.13	0.51	1,182	772	604	410	Guildford	28.93	0.47	1,242	456	452	786	Newcastle-upon-Tyne	18.87	0.52	23,217	14,590	10,864	8,627	Tenterden	29.82	0.57	1,224	771	610	453
Bridgwater	16.21	0.48	2,040	1,290	1,106	750	Halifax	18.65	0.58	7,693	5,705	5,372	1,988	Newport	23.52	0.43	1,420	625	386	795	Tewkesbury	20.98	0.54	1,189	676	431	513
Bridport	16.71	0.48	1,265	729	554	536	Harwich	25.00	0.47	705	254	224	451	Northallerton	31.84	0.57	1,693	511	222	582	Thirsk	32.20	0.59	1,197	694	546	385
Brighton	48.56	0.48	11,394	2,567	2,319	8,627	Hastings	66.13	0.47	2,757	488	479	2,269	Northampton	18.40	0.43	6,372	5,159	3,936	1,213	Thirsk	35.09	0.59	1,374	960	560	414
Bristol	29.18	0.48	28,383	8,321	7,913	15,962	Helston	15.67	0.58	1,428	1,044	322	384	Norwich	39.27	0.47	6,059	1,466	1,336	4,593	Tower Hamlets	31.21	0.37	88,736	13,579	13,211	75,157
Buckingham	3.14	0.52	1,440	1,053	403	387	Hereford	31.09	0.48	2,460	1,284	1,263	1,176	Nottingham	18.72	0.44	15,422	10,703	9,518	4,919	Truro	15.16	0.5	2,328	1,596	1,205	732
Bury	20.22	0.49	6,597	4,016	4,463	1,581	Hertford	29.64	0.49	1,131	549	541	582	Oldham	29.66	0.44	15,106	11,044	10,642	3,942	Tynemouth	17.04	0.5	5,301	4,137	2,226	1,164
Bury St. Edmunds	22.31	0.5	2,237	1,496	1,304	741	Horsham	27.32	0.36	1,175	770	672	405	Oxford	29.56	0.45	4,436	1,407	1,615	2,799	Uckfield	18.61	0.48	4,375	3,231	2,526	1,364
Calne	27.71	0.6	910	720	413	190	Huddersfield	26.40	0.58	6,508	4,406	3,368	2,102	Pembroke	19.74	0.48	3,442	2,187	1,549	1,255	Wallingford	37.42	0.58	1,465	1,052	519	303
Cambridge	34.05	0.5	3,932	1,952	1,782	1,979	Huntingdon	29.54	0.36	1,177	761	565	416	Penrith and Alston Moh	19.48	0.48	2,123	857	558	1,266	Walsall	14.89	0.41	7,362	6,048	5,207	1,161
Canterbury	19.58	0.48	5,050	2,301	2,305	1,249	Hythe	31.48	0.52	2,273	1,370	1,315	1,203	Peterborough	28.56	0.46	3,104	1,425	1,321	678	Wantage	24.04	0.57	1,159	815	378	344
Cardiff	37.95	0.44	4,995	1,540	1,307	3,425	Ipswich	19.88	0.5	7,240	4,980	4,303	2,260	Petersfield	38.33	0.62	1,040	690	361	350	Warrington	16.68	0.44	5,301	4,130	3,509	1,171
Carlisle	16.17	0.47	3,904	3,334	3,017	570	Ivybridge	19.00	0.48	2,254	1,798	797	456	Plymouth	44.62	0.55	3,907	1,207	1,205	2,695	Warwick	23.27	0.49	2,264	1,558	1,414	706
Carlisle	16.69	0.5	4,924	3,977	2,452	947	Kidderminster	15.04	0.47	3,093	2,367	1,369	726	Pontypool	17.22	0.56	2,481	1,864	361	617	Wellington	23.98	0.59	3,386	2,658	1,629	550
Carmarthen	22.46	0.48	3,023	2,044	1,461	979	King's Lynn	26.44	0.48	2,139	1,272	1,015	867	Poole	16.10	0.44	1,783	1,167	783	616	Wellington	20.59	0.58	1,364	1,138	610	326
Chatham	14.24	0.4	5,559	2,742	2,466	2,817	Kingston-upon-Hull	20.90	0.41	19,209	13,198	11,548	7,119	Portsmouth	19.78	0.4	15,214	9,385	8,859	5,849	Westminster	136.97	0.31	20,152	1,036	1,033	19,116
Cheltenham	42.07	0.52	5,450	1,995	1,956	3,495	Kirkintilloch	13.24	0.49	1,045	731	244	314	Preston	19.81	0.41	13,319	9,782	8,956	3,537	Weymouth	30.02	0.46	1,916	402	430	1,084
Chester	22.49	0.52	5,074	3,462	2,812	1,612	Lambeth	31.82	0.31	91,707	4,224	4,171	86,963	Radnor	35.26	0.59	1,038	614	368	424	Whitby	20.76	0.55	2,300	1,528	839	852
Chichester	22.20	0.48	1,203	805	755	548	Launceston	33.24	0.6	853	478	165	555	Reading	22.50	0.51	5,010	2,323	1,472	2,687	Whitehaven	13.01	0.42	4,152	3,205	1,825	947
Chippenham	33.86	0.55	1,212	747	569	465	Leeds	16.77	0.51	46,215	34,978	24,537	9,237	Retford, East	28.54	0.62	8,705	6,104	3,575	2,701	Wigan	1.45	0.44	6,412	5,130	4,870	1,302
Christchurch	54.50	0.58	1,081	564	422	517	Leicester	18.65	0.46	36,127	9,758	8,498	4,369	Richmond	31.33	0.57	813	494	316	319	Wilton	3.83	0.61	1,616	1,197	491	369
Cirencester	26.65	0.51	1,095	659	464	436	Lewisham	39.04	0.31	991	592	530	399	Ripon	16.21	0.53	1,480	1,053	611	427	Winchester	34.36	0.48	1,945	767	770	1,178
Clitheroe	23.95	0.57	1,577	1,457	1,181	540	Lincoln	23.58	0.47	1,641	863	814	778	Rochdale	19.09	0.49	7,400	5,362	4,944	1,858	Windsor	34.85	0.45	1,283	352	326	931
Cockermouth	21.40	0.57	1,400	992	560	408	Lichfield	22.43	0.52	1,252	948	688	304	Rochester	22.33	0.45	2,856	1,379	1,338	1,477	Wolverhampton	16.89	0.42	21,746	21,501	17,431	5,465
Colchester	23.95	0.55	3,689	2,450	1,916	1,239	Lincoln	21.63	0.48	4,062	3,768	2,365	1,204	Rye	38.59	0.54	1,256	1,141	388	345	Woodstock	36.16	0.6	1,188	980	511	300
Coventry	17.98	0.42	7,739	5,619	5,327	2,120	Liskeard	26.51	0.52	1,021	553	458	468	Salford	21.98	0.49	18,276	13,213	12,046	5,063	Worcester	21.55	0.47	5,808	3,118	2,960	2,766
Cricklade	44.54	0.66	7,003	4,042	1,467	2,161	Liverpool	43.08	0.48	60,553	20,458	20,397	40,059	Salisbury	22.22	0.47	1,960	1,167	1,038	813	Wycombe	19.17	0.48	1,180	1,052	871	318
Denbigh	16.32	0.36	3,454	2,460	1,323	994	London	104.89	0.28	11,784	132	132	11,482	Sandwich	21.75	0.43	2,193	1,210	1,139	983	York	20.42	0.51	8,463	4,084	3,989	3,379

CHAPTER 2

"The Years of Lead": Terrorism and Voting Behavior in the Italian 1970s

Andrea Marcucci

University of Lausanne

Abstract

This paper investigates the impact of political violence on voting patterns during the turbulent Italian 1970s. Using newly assembled pooling station-level data, I employ a difference-in-differences strategy to estimate how terrorism influences political preferences at the local level. Violent events increase the vote margin between the Christian Democracy (DC) and the Communist Party (PCI), generating a "rally around the flag" effect. Each additional attack widens the vote margin by 0.23 percentage points, doubling the gap in votes between the incumbent (DC) and the opposition party (PCI). Notably, far-right wing terrorism, historically perceived as intended to induce a conservative shift, has a more substantial impact compared to other forms of violence. The effect of terrorism is persistent, affecting voting behavior for at least three consecutive electoral terms (approximately 12 years). These results suggest that political violence can increase support for the incumbent party by creating a climate of fear. In light of the increasing levels of political polarization and violence in Western countries, this study demonstrates that a negative feedback loop between these phenomena does not necessarily exist.

Keywords: Terrorism, political violence, voting, elections, polarization.

JEL: D72, D74, F52, K42.

1. Introduction

The Capitol Hill assault on 6, January 2021, made clear how politically grounded terrorism and violence are becoming a growing issue in Western democracies. Looking at Europe, it has been estimated that among the 232 terrorist attacks that happened in Germany from 2010 to 2020, more than 50% can be attributed to extreme right-wing groups (Sabet, Liebald and Friebel, 2023). In light of the rise in polarization that democracies are facing (Gentzkow, 2016; Iyengar, Lelkes, Levendusky, Malhotra and Westwood, 2019), it is important to understand the impact of terrorism on democratic institutions. After 9/11, a growing literature has tried to estimate whether terrorism is effective in achieving its political goals (Gould and Klor, 2010; Godefroidt, 2023). While most studies focus mainly on Islamist terrorism or developing countries, there is less evidence on the impact of non-Islamist terrorism in Western democracies (Godefroidt, 2023). Nevertheless, different institutional backgrounds and perpetrator ideologies may lead to different outcomes.

This paper studies the impact of terrorism on political preferences in the context of the Italian 1970s. During this decade, Italy experienced hundreds of terrorist attacks from both right and left-wing extremist groups. It has been estimated that in the so-called "Years of Lead",[1] about 1600 people died or were severely injured as a consequence of terrorist events (ISODARCO, 1992). Focusing on the Italian context from a historical perspective has three main advantages. Firstly, by digitizing new disaggregated voting data and terrorist events, I can estimate the impact of terrorism in a multi-party Western democracy with a credible identification strategy. Secondly, the Italian case is particularly relevant given the very high level of political violence and the existence of both extreme right and left-wing terrorist groups. Finally, examining this research question in a historical setting allows for estimating the impact on political polarization without the influence of the so-called "echo chambers" (Zhuravskaya, Petrova and Enikolopov, 2020), a typical phenomenon in social media that tends to reinforce individuals' pre-existing beliefs.

In order to evaluate the effect of terrorism on individuals' political preferences, I leverage newly assembled voting data at the precinct level for the three major Italian municipalities of Milan, Rome, and Turin.[2] Specifically, I use a difference-in-differences approach to compare precincts that were affected by the attacks with those that did not experience violence before and after the terrorist events occurred. Having within-municipality variation for voting preferences and spatial coordinates of the terrorist attacks, I can compare treated and control

[1] The 1970s are often referred to as the "Years of Lead" because of the very high level of political violence perpetrated through gun shootings and bombings by far-left and far-right wing terrorist groups.

[2] These cities witnessed approximately 65% of the terrorist incidents recorded during the period under study in the Global Terrorism Database (GTD 2022).

precincts located in the same urban area. This is extremely important in a context like Italy, where political violence was by far concentrated in the major cities.

In the baseline specification, each additional political violence event widens the vote margin between the incumbent party (Christian Democracy) and the main opposition party (Italian Communist Party) by 0.23 percentage points, nearly doubling the average vote gap between the two. This impact is driven by an increase in votes for the incumbent party rather than a decrease in the consensus for the Communist Party, signaling a "rally around the flag" effect of terror.[3] The impact is more pronounced when considering those terrorist attacks organized by extreme right-wing groups with the specific purpose of generating a climate of fear and tension in the population. In precincts affected by this more relevant type of attack, there is an increase in votes for the Christian Democracy of about two percentage points caused by an additional terrorist event. Moreover, the effect is long-lasting, impacting the next three election rounds following an attack. These results seem to be driven mostly by the security concerns of the population. Indeed, using data from the Manifesto project (Lehmann et al 2023), I show that, in precincts more affected by the attacks, individuals vote for those parties with a more repressive internal security agenda.

The main caveat of the identification strategy is that it estimates only a localized impact of political violence. Indeed, exploiting within-municipality variation comes at the cost of not being able to investigate other possible channels than direct exposure to violence. Although these alternative mechanisms, such as media penetration, might be sizable, the main focus of this project is to understand the impact of a first-hand experience of violence. Especially in a context where terrorist events of different intensities were happening on a regular basis, this question is first order.

This paper contributes to several strands of research. Foremost, I add to the literature studying the impact of terrorism on political preferences. These studies fall into two broad categories: one focusing on the impact of terrorism within single countries or specific attacks, and another adopting an international, cross-country perspective. The first group is more represented in the economics literature and investigates the consequences of terrorism in various settings, sometimes reaching contradictory conclusions. While Gould and Klor (2010) find that terrorism works in increasing support for Palestinian concessions in Israel, Berrebi and Klor (2008) and Getmansky and Zeitzoff (2014) uncover an increase in vote shares for right-wing parties

[3] Although the Communist Party received on average slightly more votes than the Christian Democracy in the three municipalities I analyze, I refer to the latter as the incumbent since the PCI never received more votes than the DC in the parliamentary elections at national level and never joined any government coalition. On the other hand, the Christian Democracy expressed all the Prime Ministers during the period under study except for the years 1981 and 1982.

caused by the terrorist threat.[4] More recently, Sabet et al. (2023) show that violence increases support for the extreme right party AfD (Alternative für Deutschland) in Germany.[5] Delving into the emotional responses to violence, Vargas, Purroy, Coy, Perilla and Prem (2023) show that fear generated by landmine explosions depresses turnout in the next elections in Colombia.

The findings in the cross-country literature are even more mixed. Gassebner, Jong-A-Pin and Mierau (2008), looking at more than 100 countries, show that terrorism increases the probability for the incumbent to be replaced. Conversely, Abrahms (2006) assesses the effectiveness of major terrorist groups recognized by the US and concludes that they rarely achieve their political objectives, concluding that "terrorism does not work". In the European context, Peri, Rees and Smith (2020) analyzing survey data from the European Social Survey (ESS), find that international terrorist attacks increase the support for nationalistic parties.

Within and cross-country studies do not reach conclusive results about the effectiveness of terrorism in achieving political goals. The purpose of this paper is to improve upon the existing literature by analyzing the impact of terrorism in a new setting, using granular voting data, in the context of a Western democracy,[6] where multiple terrorist events happened over time. In particular, this study advances the existing literature in three key aspects. Firstly, considering that the emergence of terrorist groups is likely to be endogenous with respect to political preferences (Gould and Klor, 2010),[7] having within-municipality variation allows me to address this fundamental issue. Specifically, having electoral data at the precinct level and multiple episodes of political violence permits to better isolate the effect of terrorist attacks from plausible confounding factors, in line with the approach taken by Vargas et al. (2023). Secondly, by utilizing actual voting data from contexts with exceptionally high turnout (on average 95%), I minimize the influence of shifts in the composition of the electorate and address the well-known challenges associated with using post-election surveys to measure voting preferences (Funk, 2016), particularly in the context of terrorism (Montalvo, 2011). Finally, having localized data enables the identification of the direct effects of violence, unaffected by the varying salience that might result from heterogeneous media access.

---

[4] In a very different context, Montalvo (2011) shows that the 11, March 2004 bombings in Madrid significantly contributed to the victory of the opposition socialist party against the incumbent, while Balcells and Torrats-Espinosa (2018) find little impact of ETA's terrorist attacks on political preferences of the Spanish population.

[5] Regarding less stable democracies, Kibris (2011) finds that a higher number of casualties suffered by security forces in the fight against PKK is associated with an increase in votes for right-wing parties, while Rehman and Vanin (2017) observe a decline in support for democratic values among those affected by violence and terrorism.

[6] As mentioned in more systematic reviews of the research about terrorism (Godefroidt, 2023), there is a lack of studies addressing domestic terrorism in Western democratic settings.

[7] Gould and Klor (2010) discuss this issue about cross-country studies, but this can be true within-country as well if the units of observation are too aggregated.

In addition to improving the identification, the Italian context allows me to expand the scope of the analysis. Most within-country studies investigate the impact of terrorist attacks carried out by a homogeneous group of perpetrators (such as Islamic terrorism, Godefroidt, 2023). Having a multitude of attacks conducted from both extreme right and left-wing terrorist groups, I can have a better understanding of possible heterogeneous effects of terror in terms of events and perpetrators' characteristics. In addition, thanks to a panel covering numerous events, I can explore the longer-term impacts on political preferences beyond the immediate effects driven by salience, having a deeper understanding of how terrorism influences political preferences over time.

A second field of research this paper contributes to is the one related to the causes and consequences of political polarization (Gentzkow, 2016; Iyengar et al., 2019; Boxell, Gentzkow and Shapiro, 2024). In particular, I try to understand whether terrorism, being an extreme form of political demonstration, can influence political polarization of the population (Caprettini, Caesmann, Voth and Yanagizawa-Drott, 2022) or generate moderation with a rally around the flag effect. Given the rise in social movements and protests in Western democracies (Gethin and Pons, 2024) and their likelihood of escalating in more violent episodes such as the assault of Capitol Hill, it is important to understand the possible impact of this kind of violence on electoral preferences.

Finally, I contribute to the historical literature on terrorism and political movements in the 1970s. Extensive descriptive studies have examined the origins and outcomes of political violence during this era, both internationally (Caiani, Della Porta and Wagemann, 2012; Della Porta, 2013; Provenzano, 2018) and within the Italian context (Porta and Tarrow, 1986; Ferraresi, 2012). My research builds on this by causally testing the impact of terrorism on the political preferences of the most directly affected individuals in the Italian population.

The remainder of this paper is structured as follows. Section 2 describes the historical context of the years of lead. Section 3 describes the data used, Section 4 lays out the identification strategy, and Section 5 presents the baseline results and a series of heterogeneous effects. Finally, Section 5 concludes. The Appendix contains further detailed explanations about data sources and construction as well as various robustness checks.

2. Background and Context

> *"The massacres had a clear political aim: to arouse, by means of the most savage provocations, an enraged popular reaction that would have justified repressive countermeasures."*
>
> Vincenzo Vinciguerra (perpetrator of Peteano massacre)

In a 1984 national poll, 36% of Italians believed that terrorism would be the historical development from the past fifty years that historians would devote the most attention to.[8] Other significant events, such as the history of fascism, the liberation and resistance at the end of the Second World War, or the economic miracle, did not even get to half as many mentions as terrorism (Drake, 1999). This was the perceived importance of terrorist phenomenon by the Italians during the "Years of Lead".

In the period with the highest terrorist activity (1976-1980), there were about 9,673 episodes of political violence, an average of five per day (ISODARCO, 1992). This season of violence began with the Piazza Fontana bombing on December 12, 1969, a devastating attack that killed 17 people and injured 90. Initially attributed to anarchist groups, later investigations revealed that the neo-fascist group Ordine Nuovo was responsible. During the early 1970s, approximately 83% of terrorist attacks were perpetrated by neo-fascist groups, but in the second half of the decade, far-left organizations, notably the Red Brigades (Brigate Rosse), became more active, leading to a predominance of left-wing violence (ISODARCO, 1992). There are several historical theories explaining the insurgence of this phenomenon, some of which are summarized in the historical Appendix A. In a nutshell, the growing strength of the blue-collar movement, combined with fears that the inclusion of the Communist Party in the government would push Italy too close to the USSR, made extreme-right conservative forces fight back using indiscriminate violence. At the same time, far-left militants criticized the Communist Party's attempt to compromise toward moderate positions in hopes of entering government.[9] This situation led many young militants to radicalize and initiate acts of violence to incite the working class toward an armed revolution.

During the "Years of Lead", Italy witnessed 14,591 incidents of political violence due to attacks from both the far-right and far-left perpetrators, resulting in 419 deaths and at least 1,181

[8] The historical context described here, as well as the one in Appendix A, draw on the accounts of Ferraresi (2012); ISODARCO (1992); Della Porta (2013, 2018); Testa (1986); Drake (1999); Della Porta (2013).

[9] This strategy proved unsuccessful, as the Communist Party never managed to join the government.

wounded (ISODARCO, 1992). In the same period, from 1968 to 1983, the country had 21 different governments under 12 presidents, with only one not affiliated with the Christian Democracy (Testa, 1986). This paper examines whether terrorist attacks, by inducing a climate of fear and tension in the population, influenced electoral support for the incumbent party, generating a "rally around the flag" effect.

3. Data

This section describes the data sources and construction process. The analysis is based on novel digitized data on precinct-level electoral results and terrorist events from 1968 to 1983 in Italy, the so-called "Years of Lead". Additional information about the data is presented in Appendix B.

Electoral results

The study uses novel digitized electoral data at the precinct level for the Italian Chamber of Deputies (Camera dei Deputati). The national elections considered are those held in the years: 1968, 1972, 1976, 1979, and 1983. The data span coincides with the period typically recognized as the Italian "Years of Lead", marked by heightened political violence and terrorism in the country. The data have been collected from the historical archive of the Italian lower chamber (Archivio storico della Camera dei Deputati), for the cities of Milan, Rome and Turin. According to the Global Terrorism Dataset (GTD 2022), these three cities accounted for about 65% of the total number of terrorist attacks during this period. In total, I have information about 6,182 unique precincts observed for five consecutive elections.[10] Given the rules conceiving the structure of the precincts in Italy, there are, on average, 552 voters for each unit of observation. Considering the very high voting turnout in this historical period, I have approximately a panel of the same individuals casting their vote in each election.[11] The high granularity of the data allows for an analysis that uses quasi-individual voting preferences from actual votes rather than from surveys. This avoids the potential biases inherent in survey data when examining political preferences (Funk, 2016).

In order to measure the impact of terror at the local level, the location of the different precincts within cities is needed. For the municipalities of Turin and Milan, I collected historical data on the location of all the polling stations and the corresponding precincts assigned to them (see Appendix B.1). On the other hand, for the municipality of Rome, I use the centroid

[10] The number of precincts for each municipality might slightly change from one election to another, so the panel is unbalanced for some units.

[11] This is only approximately true. Indeed, for each new election, some individuals will move in and out of the sample of voters for a given precinct. In particular, people dying or changing residence will move out, while individuals coming of age or moving into the neighborhood will move in.

of the precincts constructed by Pinto (2023). This comes at the cost of introducing some noise in the estimates. The main results are robust to removing the city of Rome from the sample.

Terrorist attacks

The main dataset used to study terrorist events, the Global Terrorism Database (GTD 2022), is at the municipality level. In order to recover information about the precise lon-lat coordinates of political violence events, I digitized data from ISODARCO (1992). This book contains a full chronology of all the main violent and non-violent political events that happened during the "Years of Lead". For each event, it has information about the political affiliation of the actor perpetrating the action, the date, and a brief description of the event, specifically what happened and where. Using Chat-Gpt4 API, I performed a named entity recognition (NER) exercise to find the addresses where the events happened. Then, I geocoded the specific addresses with the Google Maps API. Additionally, I extracted information about the nature of the event, violent or not, the number of people killed, and whether or not someone was wounded.[12] In Figure A2 is reported the number of events happening each month according to the political ideology of the perpetrators. Two aspects are particularly relevant. Firstly, the attacks do not seem to happen strategically around election dates. Secondly, it is possible to observe a pattern linked to the political ideology of the terrorists. In the first half of the period, there is a preponderance of extreme right-wing attacks, while in the second half, there is a sort of reaction from extreme left-wing groups. As we can observe from Figure A3, Rome is the city where the highest number of events happened, followed by Milan and Turin. This is in line with the dimensions of the three cities and the fact that in Turin, the presence of extreme right-wing groups was very limited.

Manifesto Project data

I use data from the Manifesto Project Database to understand the evolution of parties stances on different policy topics. This dataset contains text analysis of the political manifestos for all the parties that got at least one seat in the parliament in a given election. In particular, I use an index indicating party support for "Law and Order" policies such as increasing resources for the police, promoting tougher attitudes in courts and giving more importance to internal security. Following Colantone and Stanig (2018), I compute two different measures at the precinct level, combining electoral outcomes and the value of the index for each party-election: the median voter index and the center of gravity. Once parties have been ranked according to the left-right score, lower-higher support for "Law and Order", the median voter score corresponds to the political stances in the party's manifesto voted by the median individual in each precinct. As such, it captures the shifts in preferences at the center of the

[12]For more details about the creation of the terror data, please refer to Appendix B.2.

electorate. On the other hand, the center of gravity is influenced by the whole distribution of political preferences. For this reason, it is more useful to understand the change in political preferences at the extremes of the distribution. In the absence of survey data at the precinct level, these two measures are useful for determining whether there is a change in preferences and public concern about internal security, in addition to shifts in votes for different parties.

4. Empirical Strategy

4.1 Estimation

I estimate the effect of terrorist attacks and political violence on voting preferences at the local level using a difference-in-differences strategy, where precincts located near the events (treated) are compared to precincts situated further away (controls). The unit of analysis is a precinct i, observed for the five consecutive elections, e, from 1968 to 1983. I have information on the universe of precincts located in the three major Italian municipalities of Milan, Rome, and Turin. Formally, I estimate the following regression equation:

$$y_{ime} = \alpha + \beta_1 Violence_{ime} + \gamma_i + \lambda_{me} + \varepsilon_{ime} \tag{1}$$

Where y_{ime} is a given measure of political preferences for the electorate in precinct i, located in municipality m, for election e. As main dependent variable of interest, I use the difference in vote share between the Christian Democracy (DC) and the Italian Communist Party (PCI). Indeed, the Christian Democracy was the party that appointed the prime minister during most of the period under study, while the Communist Party was the main opposition party. Subsequently to an attack, an increase in the gap between the two parties represents a "rally around the flag", with the population affected by political violence increasing the support for the incumbent party in power. On the other hand, a decrease in the vote margin determines an opposite polarizing effect of terrorism. To measure political violence and terrorism, $Violence_{ime}$, I use as main explanatory variable the number of events happening within 500 meters from a pooling station,[13] before a given election e, but after the previous election $e-1$. Alternatively, I also employ a binary variable taking value one if at least one attack happened before election e but subsequently to election $e-1$. Importantly, including in the regressions municipality-election fixed effects, λ_{me}, I can control for all time-varying confounding factors at municipality level that might be correlated with both political preferences and the likelihood of terrorist attacks.

The main treatment of interest is not an absorbing staggered treatment, taking multiple dis-

[13] For Rome, I use the number of attacks happening within 500 meters from the centroid of the precinct. In Section 4.2, I validate the radius choice, while in Appendix D.1, I show robustness for different distance values from the attacks.

crete values and switching on and off over time. For this reason, in addition to the standard regression model that incorporates precinct γ_i and municipality-election λ_{me} fixed effects, I estimate event studies in Section 5.3 using the methodology developed by De Chaisemartin and d'Haultfoeuille (2024). These event studies not only corroborate the parallel trend assumption but also demonstrate that the effects observed in the two-way fixed effects regressions cannot be solely attributed to negative weights.

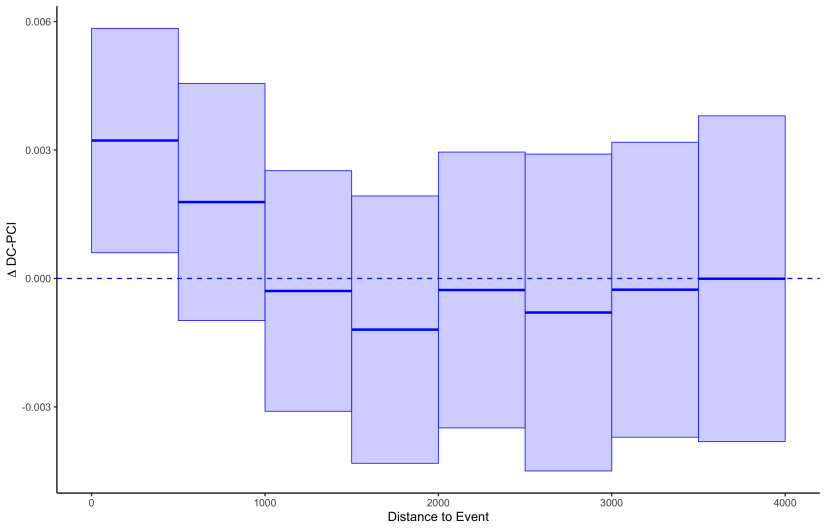
4.2 Radius Choice

In order to estimate equation 1, it is important to evaluate the spatial scope of violence. It is not straightforward to assess until what distance from violent events one might expect to observe an impact on political preferences. Previous literature primarily relies on more aggregated observational units or rural contexts (Vargas et al., 2023), which either precludes estimating the spatial scope or constitutes a very different setting, making the comparison difficult. Furthermore, this period was characterized by a very high general level of politically motivated violence. The reference point of the population regarding terror was probably very different from today's Western European countries. For the above-mentioned reasons, I estimate the spatial scope of violence in a data-driven way using the following regression model:

$$y_{ime} = \alpha + \sum_{j=0.5}^{4} \beta_j Violence_{jime} + \gamma_i + \lambda_{me} + \varepsilon_{ime} \tag{2}$$

Where I construct separate dummy variables, $Violence_{jime}$, taking value one if a given precinct i is located within the distance bin j of a violent event happening before election e. Each bin is 0.5 Km wide. Figure 1 reports the coefficients estimated in equation 3. The impact of political violence on electoral preferences is sizable and significant up to 0.5 Km from terrorist events. For this reason, in all the baseline specifications, I consider as treated all those precincts whose polling stations are located within 500 meters of violent events happening before a given election date. In Appendix D.1 results using alternative radiuses are reported as robustness.

Figure 1: Definition of treated region



Notes: The graph reports the coefficients β_j estimated with equation 3. The coefficients represent the impact for a given precinct of experiencing at least one terrorist attack at different distance bins of 500 meters as detailed in Section 4.2.

5. Main Results

5.1 Baseline Estimates

Table 1 displays the main results related to the difference in vote shares between the Christian Democracy and the Italian Communist Party ($\Delta DC - PCI$). Columns (1) to (3) report the impact of experiencing at least one terrorist attack happening before an election e. On the other hand, columns (4) to (6) display the effect of the total number of attacks happening before a given election. In columns (1) and (4), I include precinct and election by municipality fixed effects, while in columns (2) and (5), a more restrictive specification where grid cells of 25 square kilometers by election fixed effects are included.[14] Finally, in columns (3) and (6), municipality time trends are accounted for. All the coefficients are positive, suggesting a rally around the flag effect caused by political violence. Considering the more demanding specifications in columns (2), (3), (5), and (6), we can notice how the actual number of attacks seems to matter more than just the occurrence of at least one event. Interpreting the size of the coefficient of the main specification in column (4), one additional act of violence increases the margin between the Christian Democracy and the Communist Party by 0.23 percentage points, which corresponds to an increase of about 190% with respect to the mean of the dependent variable.

[14] This additional set of fixed effects allows us to control for possible within-municipality time-varying confounders that might co-determine political preferences and violence.

Table 1: Political violence impact on Delta Share

	Δ DC-PCI					
	(1)	(2)	(3)	(4)	(5)	(6)
I(attack)	0.0031**	0.0016	0.0016			
	(0.0013)	(0.0014)	(0.0014)			
# Attacks				0.0023***	0.0016***	0.0016***
				(0.0006)	(0.0006)	(0.0006)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Municipality-Election	✓			✓		
Grid 25Km2-Election		✓	✓		✓	✓
Municipality			✓			✓
Time Trends:						
Election × Municipality			✓			✓
Dep. Var. Mean	-0.0012	-0.0012	-0.0012	-0.0012	-0.0012	-0.0012
R^2	0.8747	0.8780	0.8780	0.8748	0.8780	0.8780
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. The dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. *I(attack)* is a binary variable taking value 1 if in the period prior to an election e at least one attack happens within 500 meters from a given precinct and 0 otherwise. *# Attacks* is the number of attacks happening before an election e within 500 meters from a given precinct. In columns (1)-(4) precinct and municipality-election fixed effects are included. In columns (2) and (5) 25 square kilometers by-election fixed effects and precincts fixed effects are included. In columns (3) and (6) municipality specific time trends are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

It is crucial to determine whether the observed increase in the vote share difference between the incumbent party and the opposition results from a rise in support for the Christian Democracy, a decrease in votes for the Communist Party, or both. Table 2 scrutinizes this mechanism. In column (1), the same estimates as Table 1 column (4) are reported as a way of comparison. In column (2), the dependent variable is the vote share for the Christian Democracy, while in column (3), it is the vote share for the Italian Communist Party. Finally, in column (4), the impact of violence on the total number of cast votes is reported.[15] For this last estimate, I use a Poisson regression model given the count-data nature of the dependent variable. The widening of the gap between the Christian Democracy and the Communists is entirely driven

[15]Unfortunately in the data it is not present the number of individuals eligible to vote. For this reason, it is not possible to look directly at turnout.

by an increase in votes for the incumbent, supporting a rally around the flag explanation for the reaction to politically motivated violence. Considering that the Communists' vote share does not significantly change following violent events, and that there is no change in the total number of voters (column 4), the DC increase in votes must be due to a decline in support for another minor party. This mechanism is analyzed in Section 5.4.

Table 2: Political violence impact on parties votes

	Δ DC-PCI	DC	PCI	Tot. Votes
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	Poisson
# Attacks	0.0023***	0.0023***	0.0000	-0.0008
	(0.0006)	(0.0004)	(0.0003)	(0.0010)
Fixed Effects:				
Precinct	✓	✓	✓	✓
Municipality-Election	✓	✓	✓	✓
Dep. Var. Mean	-0.0012	0.2909	0.2921	552.4205
R^2	0.8748	0.7428	0.8965	
Precincts (clusters)	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In column (1) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (3) the dependent variable is the votes' share for the Christian Democracy and the Communist Party respectively. In column (4) a Poisson model for the total number of voters is estimated. *# Attacks* is the number of attacks happening before an election e within 500 meters from a given precinct. In columns (1) to (4) precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Heterogeneous Effects

In Section 5.1, all different types of violence have been considered together, regardless of the intensity and the political ideology of the perpetrators of the attacks. To have a better understanding of possible heterogeneous effects, in terms of the political ideology of the perpetrators, Table 3 distinguishes between right and left-wing terrorist attacks. The variable *# Terror Right* counts the number of terrorist events attributable to the so-called "strategy of

tension" happening before a given election, while the variable *# Terror Left* corresponds to the number of attacks conducted by extreme left terrorist groups.[16]

Table 3: Terrorism impact on parties votes

	Δ DC-PCI	DC	PCI	Δ DC-PCI	DC	PCI
	(1)	(2)	(3)	(4)	(5)	(6)
# Terror Right	0.0190***	0.0220***	0.0029			
	(0.0056)	(0.0046)	(0.0030)			
# Terror Left				0.0015	0.0023***	0.0008
				(0.0010)	(0.0007)	(0.0006)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Municipality-Election	✓	✓	✓	✓	✓	✓
Dep. Var. Mean	-0.0012	0.2909	0.2921	-0.0012	0.2909	0.2921
R^2	0.8749	0.7434	0.8965	0.8747	0.7425	0.8965
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (4) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (5) the dependent variable is the votes' share for the Christian Democracy, while in columns (3) and (6) the dependent variable is the votes' share for Communist Party respectively. # Terror Right is the number of terrorist attacks conducted by far-right groups happening before an election e within 500 meters from a given precinct. # Terror Left is the number of terrorist attacks conducted by extreme left-wing groups happening before an election e within 500 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The general pattern we can observe is very similar to the one outlined in Table 2. Only the Christian Democracy records a significant change in votes. There is a striking difference in terms of magnitude between left and right-wing terrorism impacts. The coefficient in column (2) is almost ten times larger than the one in column (5). This is consistent with the historical narratives outlined in Section 2. The terrorist attacks which were intended to bring about a conservative shift in public opinion, actually seem to have had a much greater effect in increasing the votes for the Christian Democracy. Table A2 in Appendix D shows similar estimates for lower-scale right and left-wing political violence. For these types of smaller-

[16] See Appendix B.2 for a more detailed description of these categories.

scale violence, we detect a lower impact on voting preferences. Notably, for these events not linked to more organized terrorist groups, violence perpetrated by individuals with left-wing ideology has a stronger effect than the one committed by far-right militants.
As an alternative measure of terrorist attacks' severity, in Table 4, I consider the number of victims killed in terrorist attacks as main explanatory variable.

Table 4: Number of victims and parties votes

	Δ DC-PCI	DC	PCI	Δ DC-PCI	DC	PCI
	(1)	(2)	(3)	(4)	(5)	(6)
I(killed)	0.0083***	0.0073***	-0.0010			
	(0.0028)	(0.0019)	(0.0017)			
# Killed				0.0057***	0.0058***	0.0002
				(0.0013)	(0.0011)	(0.0008)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Municipality-Election	✓	✓	✓	✓	✓	✓
Dep. Var. Mean	-0.0012	0.2909	0.2921	-0.0012	0.2909	0.2921
R^2	0.8748	0.7425	0.8965	0.8749	0.7435	0.8965
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

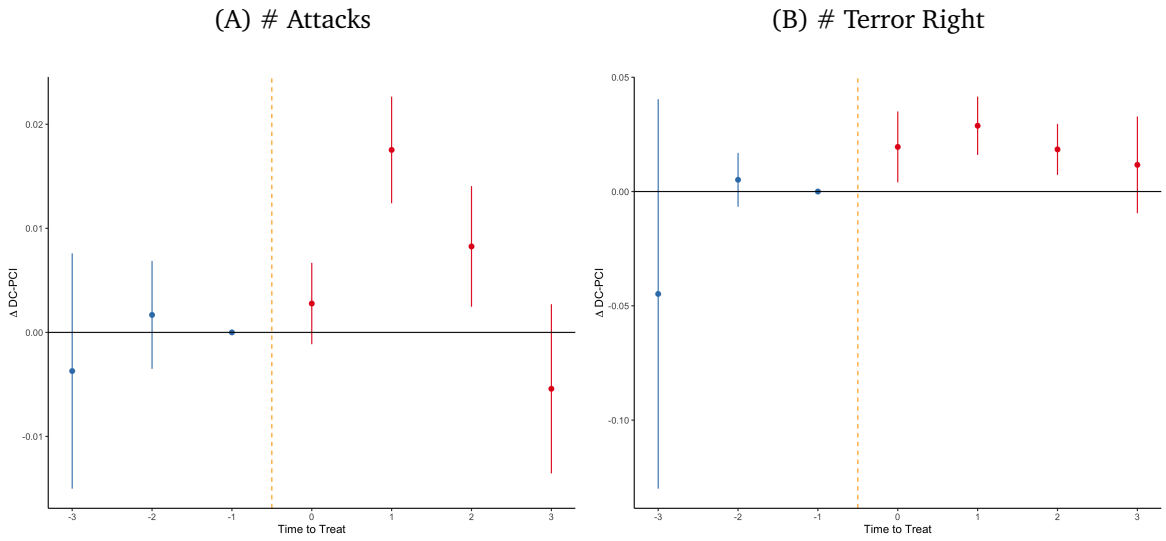
Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (4) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (5) the dependent variable is the votes' share for the Christian Democracy, while in columns (3) and (6) the dependent variable is the votes' share for Communist Party respectively. I(killed) is a binary variable taking value 1 if in the period prior to an election e at least one individual is killed in a terrorist attack within 500 meters from a given precinct and 0 otherwise. # Killed is the number of individuals getting killed in terrorist attacks happening before an election e within 500 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Compared to Table 2 it is possible to notice that taking into account deadly terrorist attacks and the number of victims the coefficients are about three times larger. Nevertheless, the impact of those attacks examined in Table 3, planned with the specific target to generate a sentiment of fear in the population, is even stronger. This suggests that those terrorist attacks conceived within the framework of the so-called "Strategy of Tension" were actually more effective in changing political preferences.

5.3 Long-Run Impact

In this section, I use an event study methodology to investigate the long-term effects of terrorism on political preferences and to examine the potential presence of pre-trends in the dependent variable. Given the nature of the treatment considered (the number of attacks), which takes multiple discrete values and switches on and off over time, the estimation of dynamic effects with a standard two-way fixed effects regression would likely be contaminated by negative weights. To solve the issue, I use the methodology developed by De Chaisemartin and d'Haultfoeuille (2024) to estimate dynamic treatment effects in a non-standard staggered setting where treatment is not absorbing.

Figure 2: Dynamic treatment effects of terrorism



NOTE: The figure displays the coefficients of estimates of leads and lags of the variables *# Attacks* in panel (a) and # Terror Right in panel (b). The estimates have been computed using the methodology developed by De Chaisemartin and d'Haultfoeuille (2024). Panel (a) contains dynamic effects similar to the static coefficient estimated in column (4) of Table 1, while panel (b) contains dynamic effects similar to the coefficient estimated in column (1) of Table 3. The dependent variable is the difference in votes between the Christian Democracy and the Communist Party Δ DC-PCI.

In Figure 2, the dynamic treatment effect plots are reported. Panel (a) shows the impact of the total number of attacks happening before an election on the vote margin between the Christian Democracy and the Communist Party, while in panel (b), I consider as explanatory variable the total number of terrorist attacks by extreme right-wing groups. In both cases, it is possible to exclude the existence of pre-trends in the outcome variable. In line with the results shown in Table 3, the impact of extreme right-wing terrorist attacks is much larger in magnitude, materializes immediately, and is more persistent. Consistently with the findings

presented in Table 3, dynamic estimates indicate that attacks designed to foster a sense of insecurity in public opinion exert a more substantial impact than other forms of violence. In general, both political violence and right-wing terrorism determine a long-lasting change in the political preferences of the voters, with an impact that remains significant for about three electoral terms (about 12 years).

5.4 Change in ideology

The change in voting preferences generated by political violence might be driven by a shift in parties' platforms. For instance, in the Israeli setting, Gould and Klor (2010) suggest that the rise in support for right-wing parties identified in previous work by Berrebi and Klor (2008), was in fact a result of the political spectrum shifting leftwards. To determine whether the surge in votes for the Christian Democracy stems from a conservative ideological shift brought about by a climate of fear and tension in the society, I utilize data on parties' political platforms from the Manifesto Project (Lehmann et al 2023). Indeed, since there is no survey on political preferences with precinct-level information for the study period, adopting a strategy similar to Gould and Klor (2010) is not possible. As detailed in Section 3, the Manifesto dataset contains different scores derived from textual analysis of the political manifestos for all the parties getting at least one candidate elected in the parliament. In particular, I focus on the importance of "Law and Order" to understand whether, in affected precincts, individuals tend to support parties with a more securitarian agenda. Following Colantone and Stanig (2018), I compute two alternative measures: the median voter party score and a center of gravity measure. While the median voter reflects how the ideology changes at the center of the voters' distribution, the center of gravity is also sensible to changes at the extremes of the political spectrum. Suppose the increase in votes for the Christian Democracy is driven by a sentiment of fear caused by terrorism. In that case, we should generally expect an increase in votes for all the parties proposing more stringent internal security measures. I look into this possible mechanism in Table 5. In particular, in columns (1) and (2), I regress the median voter and center of gravity scores for the presence of law and order policies in parties' political manifestos on the total number of violent episodes *# Attacks*, while in columns (3) and (4) I use as main explanatory variable the total number of terrorist attacks perpetrated by extreme right-wing groups, *# Terror Right*. Because of data quality reasons, detailed in Appendix B, for the analysis conducted in this section (Tables 5 and 6), the sample is restricted to those precinct-elections for which the data of all parties are precise enough. This is the cause of a reduced sample size in Tables 5 and 6. As it is possible to observe in columns (1) to (3) of Table 6, the baseline results are almost unaffected in this smaller sample.

Table 5: Shift in preferences for Law and Order

	Median Law-Order	Gravity Law-Order	Median Law-Order	Gravity Law-Order
	(1)	(2)	(3)	(4)
# Attacks	0.0084**	0.0016*		
	(0.0033)	(0.0010)		
# Terror Right			0.0785***	0.0145**
			(0.0228)	(0.0068)
Fixed Effects:				
Precinct	✓	✓	✓	✓
Municipality-Election	✓	✓	✓	✓
Dependent variable mean	0.1806	0.2243	0.1806	0.2243
R^2	0.7977	0.9722	0.7978	0.9722
Precincts (clusters)	6,120	6,120	6,120	6,120
Observations	23,303	23,303	23,303	23,303

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (3) the dependent variable is the score related to support for tougher law and order measures for the party voted by the median voter in a given precinct. In columns (2) and (4) the dependent variable is the center of gravity (Colantone and Stanig, 2018) for the law and order score, computed as detailed in Section 3. *# Attacks* is the number of attacks happening before an election e within 500 meters from a given precinct. # Terror Right is the number of terrorist attacks conducted by far-right groups happening before an election e within 500 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Interestingly, most of the impact of terrorist attacks seems to be localized at the center of voters' ideology distribution. Indeed, the effects on the center of gravity measure for law and order (columns 2 and 4) are about five times smaller than the coefficients corresponding to the impact on the median voter score (columns 1 and 3).[17] The preferences of the individuals in the center of the distribution have a considerable shift towards tougher measures in terms of internal security. The median voter's leaning towards a stricter securitarian agenda is remarkable, especially for the number of terrorist attacks conducted by extreme right-wing groups. In particular, an additional attack generates an increase in the score for law and order of the median voter of about 43% of the dependent variable mean. Table 5 suggests that the rally around the flag effect observed in the previous tables is driven by internal security concerns of the voters located mostly at the center of the political preferences' distribution.

[17]Notice that the mean of the two dependent variables Median Law-Order and Gravity Law-Order are similar in size, then we can compare the coefficients.

At the same time, this is not the case for individuals voting for more extreme parties. This is in line with the historical context outlined in Section 2. Indeed, individuals at the extremes of the voting preferences distribution, despite being more likely targeted by political violence, had very strong ideological convictions and were less likely to shift in response to terror.

In light of this result, it is important to understand the origin of the additional votes for the Christian Democracy. The increase in consensus might come from newly mobilized voters or former supporters of other parties. In particular, following the results about the median voter's shift, we should expect a decrease in votes for some other parties at the center of the political arena. Voter mobilization is a less probable channel given the extremely high voting turnout (on average 95%) existing in that context.

In Table 6, I report the change in vote share for all the major parties included in the Manifesto Project Database[18] and a Poisson model for the total number of voters in each precinct.[19] Panel A displays the results using the number of violent events *# Attacks* as main explanatory variable, while in Panel B reports estimates for the number of terrorist events related to far-right groups, *# Terror Right*. The increase in votes for the Christian Democracy does not seem to be driven by a stronger voter mobilization in affected precincts. If anything, the number of voters slightly decreases. Looking at the vote share received by the various parties, the increase in votes for the Christian Democracy is mirrored by an analogous decrease for the Italian Liberal Party (PLI). This is in line with the results shown in Table 5. Indeed, we can observe a shift in preferences from a center-right party, the liberals, to the major incumbent political force, the Christian Democracy. At the same time, the transfer of votes from more extreme parties to the Christian Democracy does not seem to happen. Especially for the case of extreme right-wing terrorism (Panel B), the loss of votes for the PLI is almost exactly the same magnitude as the gain of votes for the DC.

[18] In 1968, the two main socialist parties, PSI and PSDI, run in a unique coalition for the elections the PSU (Unified Socialist Party). Given that for all the other elections in the sample, the two parties run separately, I reassign the votes of PSU to PSI and PSDI proportionally with respect to the vote shares that the two parties got in the subsequent elections of 1972. For this reason, the coefficients of these two parties should be taken with caution.

[19] As previously mentioned, the voting turnout is not present in the data. For this reason, I use the total number of voters in each precinct as a proxy for turnout. Given that this is a count variable, I estimate the impact of the attacks with a Poisson model instead of OLS.

Table 6: Impact for all parties and total voters

	Δ DC-PCI	DC	PCI	PRI	PLI	PSDI	PSI	DEMPROL	PRAD	MSI.DN	Tot. Voters
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	Poisson
Panel A:											
# Attacks	0.0020***	0.0020***	0.0000	0.0006***	-0.0031***	-0.0002*	-0.0002	0.0001	0.0000	0.0000	-0.0037***
	(0.0006)	(0.0004)	(0.0003)	(0.0001)	(0.0004)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0012)
Panel B:											
# Terror Right	0.0233***	0.0216***	-0.0017	0.0019*	-0.0221***	0.0004	-0.0007	-0.0008	-0.0003	-0.0039***	-0.0106
	(0.0057)	(0.0045)	(0.0018)	(0.0010)	(0.0037)	(0.0007)	(0.0014)	(0.0018)	(0.0011)	(0.0015)	(0.0071)
Fixed Effects:											
Precinct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Municipality-Election	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Dependent variable mean	0.0005	0.2983	0.2978	0.0440	0.0656	0.0461	0.0912	0.0194	0.0382	0.0918	553.0218
R^2	0.9067	0.8115	0.9360	0.8122	0.8247	0.7241	0.7770	0.8152	0.8531	0.8669	
Precincts (clusters)	6,120	6,120	6,120	6,120	6,120	6,120	6,120	5,512	6,008	6,120	6,128
Observations	23,303	23,303	23,303	23,303	23,303	23,295	23,295	8,578	13,827	23,303	23,355

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In column (1) the dependent variable is the vote margin between the Christian Democracy and the Communist Party. In columns (2) to (10) the vote share for various parties in the following order are reported: Christian Democracy, Italian Communist Party, Italian Republican Party, Italian Liberal Party, Italian Democratic Socialist Party, Italian Socialist Party, Proletarian Democracy, Radical Party, Italian Social Movement-National Democracy. In column (11) the dependent variable is the total number of voters. *# Attacks* is the number of attacks happening before an election e within 500 meters from a given precinct. # Terror Right is the number of terrorist attacks conducted by far-right groups happening before an election e within 500 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6. Conclusion

Does terrorism influence political preferences by increasing polarization, or does it generate a "rally around the flag" effect? I investigate this question in the context of the so-called "Years of Lead", a decade in which Italy experienced a spike in political violence and terrorism (1968-1983). By constructing a novel electoral dataset at the precinct level for the cities of Milan, Rome and Turin, and a panel of geocoded violent events, I estimate the impact of terror with a difference-in-differences strategy. Exploiting within-municipality variation, I can control for confounding factors that make violence endogenous with respect to political preferences at the local level.

The baseline estimates indicate that an additional act of political violence increases the vote margin between the incumbent, Christian Democracy, and the Communist Party by 0.23 percentage points, almost doubling the gap between the two parties. The effect is stronger and more persistent when considering terrorist attacks perpetrated by far-right groups. This is

consistent with historical narratives that depict far-right terrorism as intended to create a climate of insecurity in the population in order to promote a conservative shift. Furthermore, taking into account the party's political platforms from the Manifesto Project (Lehmann et al 2023), it is possible to conclude that precincts more exposed to terror vote for those parties proposing stricter security measures.

The findings suggest that terrorism and political violence can indeed lead the electorate to vote for the incumbent party by generating a climate of fear and insecurity, thus triggering a "rally around the flag" effect. This is in line with the literature in political psychology suggesting that it is anger pushing the electorate to vote for extreme parties rather than the fear generated by violence (Vasilopoulos, Marcus, Valentino and Foucault, 2019). Considering this result in light of the increasing levels of political polarization and violence in Western countries, it is possible to conclude that a negative feedback loop between the two does not necessarily exist. On the other hand, high levels of violence may indeed reduce polarization, leading the electorate to vote for more moderate forces that are viewed as a lifeline against the fear generated by terrorism.

While conducting the analysis at the local level in a context absent of social media allows for a clean estimation of the direct impact of violence on political preferences, other general equilibrium effects might be at play in modern societies. For example, Hatte, Madinier and Zhuravskaya (2021) show that the use of social media leads even standard news outlets to broadcast more emotionally charged content. This might potentially amplify the impact of terror on a greater number of voters. My results suggest that further research on the relationship between terrorism and political preferences, possibly using granular voting data from various contexts and countries, is essential to better understand these complex dynamics.

Bibliography

Abrahms, M. (2006), 'Why terrorism does not work', *International Security* **31**(2), 42–78.

Balcells, L. and Torrats-Espinosa, G. (2018), 'Using a natural experiment to estimate the electoral consequences of terrorist attacks', *Proceedings of the National Academy of Sciences* **115**(42), 10624–10629.

Berrebi, C. and Klor, E. F. (2008), 'Are voters sensitive to terrorism? direct evidence from the israeli electorate', *American Political Science Review* **102**(3), 279–301.

Boxell, L., Gentzkow, M. and Shapiro, J. M. (2024), 'Cross-country trends in affective polarization', *Review of Economics and Statistics* **106**(2), 557–565.

Caiani, M., Della Porta, D. and Wagemann, C. (2012), *Mobilizing on the extreme right: Germany, Italy, and the United States*, Oxford University Press.

Caprettini, B., Caesmann, M., Voth, H.-J. and Yanagizawa-Drott, D. (2022), 'Going viral: Protests and polarization in 1932 hamburg'.

Colantone, I. and Stanig, P. (2018), 'The trade origins of economic nationalism: Import competition and voting behavior in western europe', *American Journal of Political Science* **62**(4), 936–953.

De Chaisemartin, C. and d'Haultfoeuille, X. (2024), 'Difference-in-differences estimators of intertemporal treatment effects', *Review of Economics and Statistics* pp. 1–45.

Della Porta, D. (2013), *Clandestine political violence*, Cambridge University Press.

Della Porta, D. (2018), 'Radicalization: A relational perspective', *Annual Review of Political Science* **21**, 461–474.

Drake, R. (1999), 'Italy in the 1960s: A legacy of terrorism and liberation', *South Central Review* **16**, 62–76.

Ferraresi, F. (2012), *Threats to democracy: the radical right in Italy after the war*, Princeton University Press.

Funk, P. (2016), 'How accurate are surveyed preferences for public policies? evidence from a unique institutional setup', *Review of Economics and Statistics* **98**(3), 442–454.

Gambi, L. (1973), *Da città ad area metropolitana*.

Gassebner, M., Jong-A-Pin, R. and Mierau, J. O. (2008), 'Terrorism and electoral accountability: One strike, you're out!', *Economics Letters* **100**(1), 126–129.

Gentzkow, M. (2016), 'Polarization in 2016', *Toulouse Network for Information Technology Whitepaper* **1**.

Gethin, A. and Pons, V. (2024), Social movements and public opinion in the united states, Technical report, National Bureau of Economic Research.

Getmansky, A. and Zeitzoff, T. (2014), 'Terrorism and voting: The effect of rocket threat on voting in israeli elections', *American Political Science Review* **108**(3), 588–604.

Godefroidt, A. (2023), 'How terrorism does (and does not) affect citizens' political attitudes: a meta-analysis', *American Journal of Political Science* **67**(1), 22–38.

Gould, E. D. and Klor, E. F. (2010), 'Does terrorism work?', *The Quarterly Journal of Economics* **125**(4), 1459–1510.

Hatte, S., Madinier, E. and Zhuravskaya, E. (2021), 'Reading twitter in the newsroom: How social media affects traditional-media reporting of conflicts'.

ISODARCO (1992), *Venti anni di violenza politica in Italia: 1969-1988: cronologia ed analisi statistica: una ricerca ISODARCO*, Università degli studi La Sapienza.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. and Westwood, S. J. (2019), 'The origins and consequences of affective polarization in the united states', *Annual review of political science* **22**, 129–146.

Kibris, A. (2011), 'Funerals and elections: The effects of terrorism on voting behavior in turkey', *Journal of Conflict Resolution* **55**(2), 220–247.

Lehmann, P., Franzmann, S., Burst, T., Regel, S., RiethmÃ¼ller, F., Volkens, A., WeÃŸels, B. and Zehnter, L. (2023), 'The manifesto data collection. manifesto project (mrg/cmp/marpor). version 2023a'.
URL: *https://doi.org/10.25522/manifesto.mpds.2023a*

Montalvo, J. G. (2011), 'Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections', *Review of Economics and Statistics* **93**(4), 1146–1154.

National Consortium for the Study of Terrorism and Responses to Terrorism (START) (2022), 'Global terrorism database 1970 - 2020 [data file]'. Accessed: insert-date-here.
URL: *https://www.start.umd.edu/gtd*

Peri, G., Rees, D. I. and Smith, B. (2020), 'Terrorism, political opinions, and election outcomes:

Evidence from europe'.

Pinto, G. (2023), 'Sezioni elettorali italiane (sei): a new database of italian electoral results geocoded at precinct level', *Political Research Exchange* **5**(1), 2185158.

Porta, D. d. and Tarrow, S. (1986), 'Unwanted children: Political violence and the cycle of protest in italy, 1966–1973', *European Journal of Political Research* **14**(5-6), 607–632.

Provenzano, L. (2018), "power is in the streets': Protest and militancy in france, italy and west germany, 1968–1979', *Contemporary European History* pp. 1–18.

Rehman, F. U. and Vanin, P. (2017), 'Terrorism risk and democratic preferences in pakistan', *Journal of Development Economics* **124**, 95–106.

Sabet, N., Liebald, M. and Friebel, G. (2023), Terrorism and voting: The rise of right-wing populism in germany, Technical report, CESifo.

Testa, G. P. (1986), *Terrorismo: la strategia che viene dall'alto*, a cura della Associazione familiari vittime strage alla stazione di Bologna ....

Vargas, J. F., Purroy, M. E., Coy, F., Perilla, S. and Prem, M. (2023), 'Fear to vote: Explosions, salience, and elections', *Salience, and Elections (July 18, 2023)* .

Vasilopoulos, P., Marcus, G. E., Valentino, N. A. and Foucault, M. (2019), 'Fear, anger, and voting for the far right: Evidence from the november 13, 2015 paris terror attacks', *Political Psychology* **40**(4), 679–704.

Zhuravskaya, E., Petrova, M. and Enikolopov, R. (2020), 'Political effects of the internet and social media', *Annual review of economics* **12**, 415–438.

Appendix

A. Appendix: Historical Background

In Italy, the decade of the 1970s is often referred to as "Years of Lead". This expression has been coined because of the extreme level of violence that characterized those years in Italy. During the period going from January 1969 to December 1987, the country witnessed 14,591 events of politically motivated violence, 8 massacres, with a total number of 419 deaths and 1,181 wounded (ISODARCO, 1992). This season of violence that swept through Italy for about fifteen years started in the aftermath of the student international movement of 1968. While student protests were more or less confined to 1968 in most Western countries, in Italy they extended into the "Hot Autumn" of 1969, fueled by blue collars' instances for better working conditions. Exactly at the end of the autumn, on 12 December 1969, the massacre of Piazza Fontana happened, marking the start of the so-called "Years of Lead". In this terrorist attack, two bombs exploded at the National Bank of Agriculture, in Milan city center, killing 17 people and injuring another 90. In the beginning, the responsibility for this attack was given to anarchist groups, while in the subsequent years, it was made clear that the bombing had been organized by a far-right group from Padua linked to the neo-fascist movement Ordine Nuovo (New Order).[20] Initially, terrorism was mainly a far right-wing phenomenon, with about 83% of the attacks from 1969 to 1975 conducted by neo-fascist groups (ISODARCO, 1992). Subsequently, extremist left-wing organizations, notably the Red Brigades (Brigate Rosse), intensified their activities, leading to a predominance of far-left terrorist attacks over those committed by right-wing factions.

There are several theories and explanations about the origins of neo-fascist terrorism in the first years of this violent season. Many of them agree about interpreting all this violence in the setting of a more comprehensive "strategy of tension", aimed at fostering a conservative shift. The economic boom that happened in the '50s had led to deep transformations of the Italian society. In particular, among other factors, the "economic miracle" had been possible thanks to the availability of a very cheap labor force coming from the countryside in the cities (Ferraresi, 2012).[21] These profound social transformations had not been accompanied by political changes, with center governments led by the Christian Democracy for the entire period that never included parties representing the working class, leaving real wages almost

[20] As for many other terrorist attacks linked to far-right groups and the strategy of tension, nobody has been sentenced with a final conviction for the massacre, but the responsibility of the bombing was attributed to some members of Ordine Nuovo (New Order), an extreme right-wing movement.

[21] It has been estimated that about 15 million people, out of a population of 50 million, moved from the countryside, and from the south, in the cities during the economic miracle, changing completely the social structure of the country (Gambi, 1973).

unaltered. These, among others, were the reasons behind the important season of protests and strikes of the "hot autumn" in 1969. The working class movements got stronger in the negotiation with the entrepreneurs, and despite the Communist Party never entering into any government, ultimately labor unions managed to obtain the Workers' Statute.[22] On top of this important reform granting several new rights to the workers, the other main progressive laws that passed, such as the one legalizing divorce and the law establishing popular referendums, were established more as a consequence of a demand for democratization coming from collective protests rather than from an action of the party system (Ferraresi, 2012). The increase in workers' strength, combined with the perceived threat that the entry of the Communists into the government would have implied a possible shift of the country toward the Soviet bloc,[23] led part of the most conservative forces in the country to fight back by using indiscriminate violence.

At the same time, the end of the economic miracle, rising unemployment and inflation, also due to international events such as the Yom Kippur War, increased social tensions, leading some extreme left activists to radicalize more in front of a political landscape unable to give appropriate responses. In particular, the Communist Party was seen as incapable of having a proper political weight and to have abandoned the socialist ideal of armed struggle and revolution. Social unrest among the working class, the crisis and the increased level of violence perpetrated by extreme right groups made far-left terrorist groups' ranks to grow and in some cases earn the connivance of part of the blue collars.[24] The most prominent left-wing terrorist group was by far the Red Brigades (Brigate Rosse), which ultimately culminated in their spiral of violence with the kidnapping and murder of the secretary of Christian Democracy and former Prime Minister Aldo Moro in the spring of 1978.

The aftermath of the great student protest movement of 1977 gave new fertile ground to terrorist groups from both the right and the left, leading to the so-called season of "armed spontaneity" (Ferraresi, 2012). Extremists who had participated in the protests pursued their revolutionary goals by joining already existing terrorist groups or forming new ones, such as the NAR (Armed Revolutionary Groups) at the extreme right and PL (First Line) at the extreme left. The surge of a multitude of groups led to a season where, in the main cities such as Milan and Rome, there was a sort of gang warfare with frequent episodes of violence to impose on the territory. This new season culminated with the most deadly massacre of the years of lead, the bombing in Bologna station, where 85 people died, perpetrated by some

[22] The Workers' Statute, or Workers Chart, was the most important achievement of the working class movement in the post-World War II period.

[23] This fear was shared at least by part of the US intelligence and government.

[24] Left-wing terrorists were sometimes defined as "comrades who make mistakes".

members of the NAR.[25]
Turning to the political arena, during the 16 years considered in this paper (from 1968 to 1983), there were 21 different governments chaired by 12 distinct Presidents. Only one of them, Giovanni Spadolini (1981-1982), was not a member of the Christian Democracy.
The pervasive and enduring climate of violence experienced by the Italian population during the years of lead sets the context for the examination of how terrorist acts influenced electoral support, potentially fostering a "rally around the flag" effect caused by a climate of widespread fear and uncertainty.

B. Appendix: Data

This appendix provides further information on the data used in the paper. Summary statistics are reported in Table A1

B.1 Polling stations and precincts locations

The number of precincts and their location change over time because of changes in the structure of cities and inhabitants. No complete map containing historical changes in the boundaries of Italian precincts exists. To overcome this issue, Pinto (2023) reconstructed the boundaries of the current precincts for seven major Italian municipalities. Given the changes that Italian cities have undertaken in the last fifty years, using the current boundaries might lead to errors in identifying the correct location of precincts. This is why I collected historical data on the addresses of polling stations for the cities of Turin and Milan. In particular, for Turin, I have data on polling stations for the year 1995[26], while for Milan I collected data about the location of polling stations for 1963, very close to the study period.[27] For the municipality of Rome, I could not find older data regarding the location of polling stations or precincts. Notice that for this city the change in terms of number of precincts is less relevant if we compare it with Turin or Milan, going from about 3000 units during the seventies to about 2500 units for the current precincts.

[25] As material perpetrators of the massacre have been convicted members of the NAR, a far-right group, while regarding the instigators it has been hypothesized that the direction of the Bologna bombing was made by members of the Masonic Lodge P2.

[26] The last election year in which there was a number of precincts in the same order of magnitude as in the seventies. Indeed, after 1997, the number of precincts for the city of Turin passed from 1552 to 919, the current number.

[27] I thank Alberto De Cristofaro and Fondazione ISEC for providing me this data.

B.2 Terrorist attacks

The Global Terrorism Database (GTD 2022)does not provide within-municipality coordinates for terrorist attacks. For this reason, I construct a new dataset with information about political violent events, digitizing the book "Venti anni di violenza political in Italia" (ISODARCO, 1992).[28] This historical book contains a complete chronology of all the main violent and non-violent political events that happened in Italy from 1969 up to 1988. For each event is reported the date, the political ideology of the perpetrators, the city where the event happened, and a short description. The events are categorized into six different groups: 1) organized left-wing terrorism, 2) organized right-wing terrorism, 3) smaller-scale left-wing violence, 4) smaller-scale right-wing violence, 5) unknown, and 6) public disorders from both right and left-wing groups. In particular, in group 1) are included those attacks that are perpetrated by proper extreme left terrorist groups, conducting clandestine activities. On the other hand, in group 2) are included all those events identified by the authors of the book as part of a broader strategy of tension typically carried out by extreme right-wing groups, sometimes with the connivance of part of public authorities. Unless differently specified, the main independent variable used in the analysis accounts for all these categories of violence except the public disorder one since this kind of violence is more related to clashes between protesters and the police. Moreover, when distinguishing between right and left-wing terrorism, as in Table 3, I use violent events from categories 1) (*# Terror Left*) and 2) (*# Terror Right*) being events from proper terrorist organizations. A page of the book is displayed in Figure A1 as an example.

After having converted the entire book into strings, I isolate the text for each event. The remaining pieces of information I need to extract are whether the event is violent, the event's location, and the number of people killed or wounded. In order to follow a scalable procedure, I use the API of Chat-Gpt4 to extract this information. The first step is filtering out the event type (violent or not). To do this, I use the following prompt "*Use your NLP capabilities to determine whether the following event is a terrorist attack-political violence or an event related to a terrorist attack-political violence but not the attack itself. Respond with 1 if the event is a terrorist attack-political violence and 0 if it is not*:". The task has been formulated in this rather complex way given the presence of many descriptions of trials related to violent events where violence is described, but it is actually not perpetrated in that location-date.

As a second step, I identify the location where an event happens. In Natural Language Processing, this is called Named Entity Recognition task (NER). In particular, I use the following prompt "*Use your NLP capabilities to extract from the following Italian text the location where the event happened. This is a named entity recognition work. Reply only with the location name*

[28] In English "Twenty years of political violence in Italy".

or report NA if there is no specific location in the text. As location I mean an address or a public place, I am not interested in just the city:". I use similar procedures to identify the number of killed individuals and whether or not some victims were wounded.[29]

In Figure A2, the monthly number of events for the two main categories of right and left-wing terrorism is reported. Consistently with the historical facts, during the first part of the sample period, there is a predominance of extreme right-wing attacks, followed by a "reaction" by left-wing groups during the second half of the years of lead. Importantly, terrorist attacks appear to be widespread over time and not happening just in the months close to the election dates. This is in line with a more general climate of violence directed to create "tension" in the population.

In Figure A3, the monthly number of events happening in each city is reported. Given their size and the fact that extreme right-wing groups were not active in Turin, Rome and Milan have a higher number of attacks.

Figure A4 displays the spatial distribution of the events in the three municipalities according to the political ideology of the perpetrators. The underlying maps are the current boundaries of the precincts for the three cities taken from Pinto (2023).[30]

B.3 Electoral data

The electoral data at precinct level have been obtained from the Historical Archive of the Italian Chamber of Deputies (Archivio Storico della Camera dei Deputati).[31] The raw data are scans of handwritten tables where the votes obtained from the different parties in each precinct have been recorded. The digitization process was carried out with a two-step procedure. Firstly, I performed optical character recognition of the images using the Microsoft Azure reader API. As a second step, I manually fixed the mistakes made by the reader for the two most important parties considered in the analysis, the Communist Party and the Christian Democracy. For this reason, results in Section 5.4 take into account only a subset of the full data available in the analysis. Indeed, in order to compute the full distribution of political preferences, I need to use the percentages obtained by all the parties. Given that manually cleaning the data for all the parties participating in the elections would require a considerable amount of time, I include in this analysis only those units of observations (precincts-election

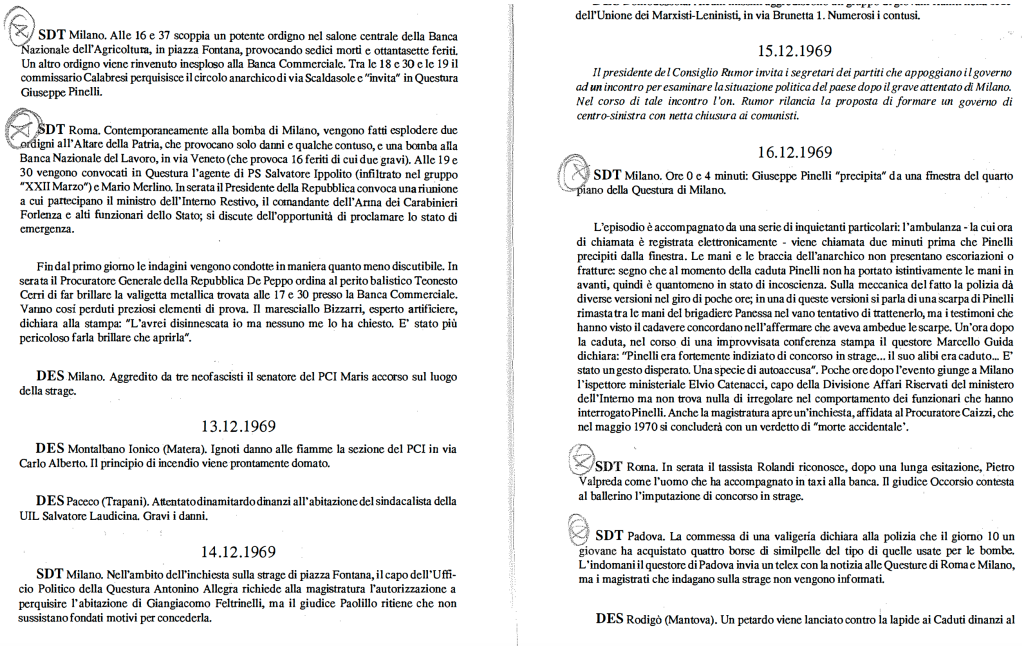
[29] Unfortunately, in most of the cases the chronology reports "multiple individuals were injured" without specifying the number.

[30] Only for the city of Rome I use the centroids of the current precincts in the analysis. As better detailed in Section B.1, for Milan and Turin I use the historical location of polling stations and I lack information for the boundaries. For consistency and to give an idea of what the precincts' boundaries look like, in these maps I report the current boundaries.

[31] I thank Giampiero Sica, the superintendent of the archive, as well as all the employees who helped with the scan of the data for their helpful work in providing me with the data at precinct-level.

years) for which the discrepancy between the total number of voters and the sum of votes received by the parties is lower than 50 votes (about 10% of the total number of voters in each precinct). This is a shortcut to consider only those precincts for which the mistakes of the OCR are not too large, in order to have conservative estimates of the effects.

Figure A1: Example Chronology



SDT Milano. Alle 16 e 37 scoppia un potente ordigno nel salone centrale della Banca Nazionale dell'Agricoltura, in piazza Fontana, provocando sedici morti e ottantasette feriti. Un altro ordigno viene rinvenuto inesploso alla Banca Commerciale. Tra le 18 e 30 e le 19 il commissario Calabresi perquisisce il circolo anarchico di via Scaldasole e "invita" in Questura Giuseppe Pinelli.

SDT Roma. Contemporaneamente alla bomba di Milano, vengono fatti esplodere due ordigni all'Altare della Patria, che provocano solo danni e qualche contuso, e una bomba alla Banca Nazionale del Lavoro, in via Veneto (che provoca 16 feriti di cui due gravi). Alle 19 e 30 vengono convocati in Questura l'agente di PS Salvatore Ippolito (infiltrato nel gruppo "XXII Marzo") e Mario Merlino. In serata il Presidente della Repubblica convoca una riunione a cui partecipano il ministro dell'Interno Restivo, il comandante dell'Arma dei Carabinieri Forlenza e alti funzionari dello Stato; si discute dell'opportunità di proclamare lo stato di emergenza.

Fin dal primo giorno le indagini vengono condotte in maniera quanto meno discutibile. In serata il Procuratore Generale della Repubblica De Peppo ordina al perito balistico Teonesto Cerri di far brillare la valigetta metallica trovata alle 17 e 30 presso la Banca Commerciale. Vanno così perduti preziosi elementi di prova. Il maresciallo Bizzarri, esperto artificiere, dichiara alla stampa: "L'avrei disinnescata io ma nessuno me lo ha chiesto. E' stato più pericoloso farla brillare che aprirla".

DES Milano. Aggredito da tre neofascisti il senatore del PCI Maris accorso sul luogo della strage.

13.12.1969

DES Montalbano Ionico (Matera). Ignoti danno alle fiamme la sezione del PCI in via Carlo Alberto. Il principio di incendio viene prontamente domato.

DES Paceco (Trapani). Attentato dinamitardo dinanzi all'abitazione del sindacalista della UIL Salvatore Laudicina. Gravi i danni.

14.12.1969

SDT Milano. Nell'ambito dell'inchiesta sulla strage di piazza Fontana, il capo dell'Ufficio Politico della Questura Antonino Allegra richiede alla magistratura l'autorizzazione a perquisire l'abitazione di Giangiacomo Feltrinelli, ma il giudice Paolillo ritiene che non sussistano fondati motivi per concederla.

dell'Unione dei Marxisti-Leninisti, in via Brunetta 1. Numerosi i contusi.

15.12.1969

Il presidente del Consiglio Rumor invita i segretari dei partiti che appoggiano il governo ad un incontro per esaminare la situazione politica del paese dopo il grave attentato di Milano. Nel corso di tale incontro l'on. Rumor rilancia la proposta di formare un governo di centro-sinistra con netta chiusura ai comunisti.

16.12.1969

SDT Milano. Ore 0 e 4 minuti: Giuseppe Pinelli "precipita" da una finestra del quarto piano della Questura di Milano.

L'episodio è accompagnato da una serie di inquietanti particolari: l'ambulanza - la cui ora di chiamata è registrata elettronicamente - viene chiamata due minuti prima che Pinelli precipiti dalla finestra. Le mani e le braccia dell'anarchico non presentano escoriazioni o fratture: segno che al momento della caduta Pinelli non ha portato istintivamente le mani in avanti, quindi è quantomeno in stato di incoscienza. Sulla meccanica del fatto la polizia dà diverse versioni nel giro di poche ore; in una di queste versioni si parla di una scarpa di Pinelli rimasta tra le mani del brigadiere Panessa nel vano tentativo di trattenerlo, ma i testimoni che hanno visto il cadavere concordano nell'affermare che aveva ambedue le scarpe. Un'ora dopo la caduta, nel corso di una improvvisata conferenza stampa il questore Marcello Guida dichiara: "Pinelli era fortemente indiziato di concorso in strage... il suo alibi era caduto... E' stato un gesto disperato. Una specie di autoaccusa". Poche ore dopo l'evento giunge a Milano l'ispettore ministeriale Elvio Catenacci, capo della Divisione Affari Riservati del ministero dell'Interno ma non trova nulla di irregolare nel comportamento dei funzionari che hanno interrogato Pinelli. Anche la magistratura apre un'inchiesta, affidata al Procuratore Caizzi, che nel maggio 1970 si concluderà con un verdetto di "morte accidentale'.

SDT Roma. In serata il tassista Rolandi riconosce, dopo una lunga esitazione, Pietro Valpreda come l'uomo che ha accompagnato in taxi alla banca. Il giudice Occorsio contesta al ballerino l'imputazione di concorso in strage.

SDT Padova. La commessa di una valigeria dichiara alla polizia che il giorno 10 un giovane ha acquistato quattro borse di similpelle del tipo di quelle usate per le bombe. L'indomani il questore di Padova invia un telex con la notizia alle Questure di Roma e Milano, ma i magistrati che indagano sulla strage non vengono informati.

DES Rodigò (Mantova). Un petardo viene lanciato contro la lapide ai Caduti dinanzi al

Notes: The bar plot displays the total number of terrorist events categorized according to the political faction of the perpetrators. Data have been digitalized by the author from ISODARCO (1992).

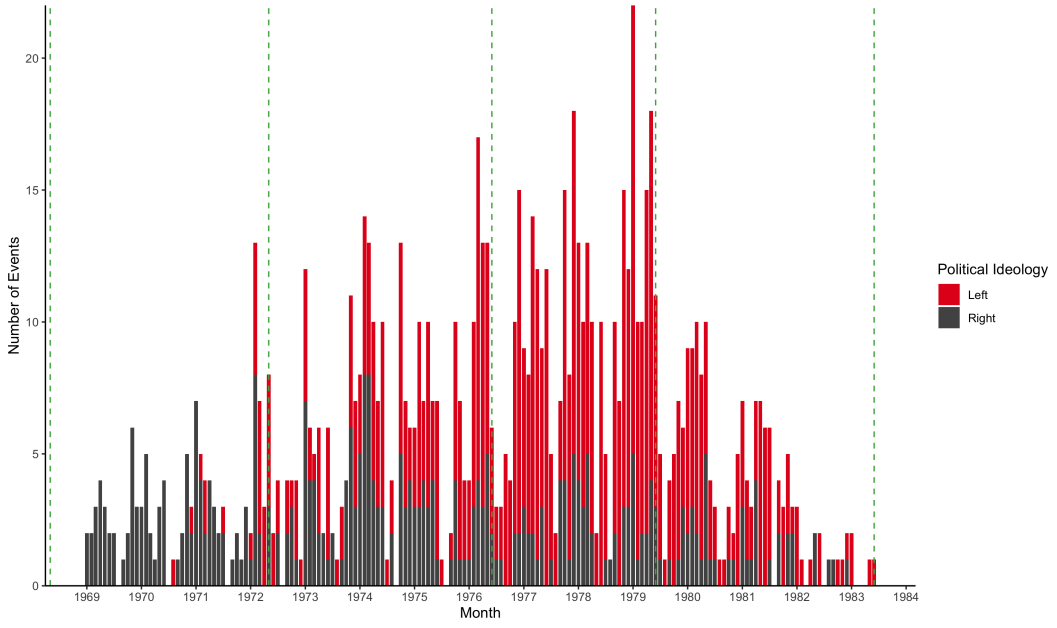
Table A1: Summary statistics

Variable	Mean	SD	Min	Median	Max	N
Panel A: *Electoral data*						
Δ DC-PCI	-0.0012	0.1812	-0.7301	0.0087	0.7694	29,661
DC	0.2909	0.0835	0	0.2869	0.7989	29,661
PCI	0.2921	0.1244	0	0.2840	0.7961	29,661
PRI	0.0440	0.0341	0	0.0348	0.4170	23,303
PLI	0.0656	0.0668	0	0.0418	0.5045	23,303
PSDI	0.0415	0.0202	0	0.0383	0.4513	18,124
PSI	0.0886	0.0319	0	0.0855	0.4553	18,124
DEMPROL	0.0194	0.0112	0	0.0177	0.1688	8,578.0
PRAD	0.0382	0.0282	0	0.0363	0.3434	13,827
MSI.DN	0.0918	0.0526	0	0.0816	0.5134	23,303
Tot. Votes	552.42	126.50	27.000	570.00	1,431.0	29,661
Median Law-Order	0.1806	0.6356	-0.6931	0.0296	1.5974	23,303
Gravity Law-Order	0.2243	0.4936	-0.6490	-0.0264	1.3934	23,303
Panel B: *Political violence data*						
# Attacks	0.4597	1.1151	0	0	13.000	29,661
I(attack)	0.2386	0.4263	0	0	1.0000	29,661
# Killed	0.0702	0.5597	0	0	17.000	29,661
I(killed)	0.0355	0.1850	0	0	1.0000	29,661
# Terror Right	0.0128	0.1426	0	0	3.0000	29,661
# Terror Left	0.1399	0.4914	0	0	7.0000	29,661
# Attacks Right	0.1404	0.4985	0	0	7.0000	29,661
# Attacks Left	0.1113	0.4266	0	0	7.0000	29,661

Notes: The table reports summary statistics for the main variables used in the analysis. The unit of observation is a precinct and election. In *Panel A*, I report summary statistics for the electoral data used as dependent variables. In *Panel B*, I report the summary statistics for the measures of political violence and terrorism used as explanatory variables.

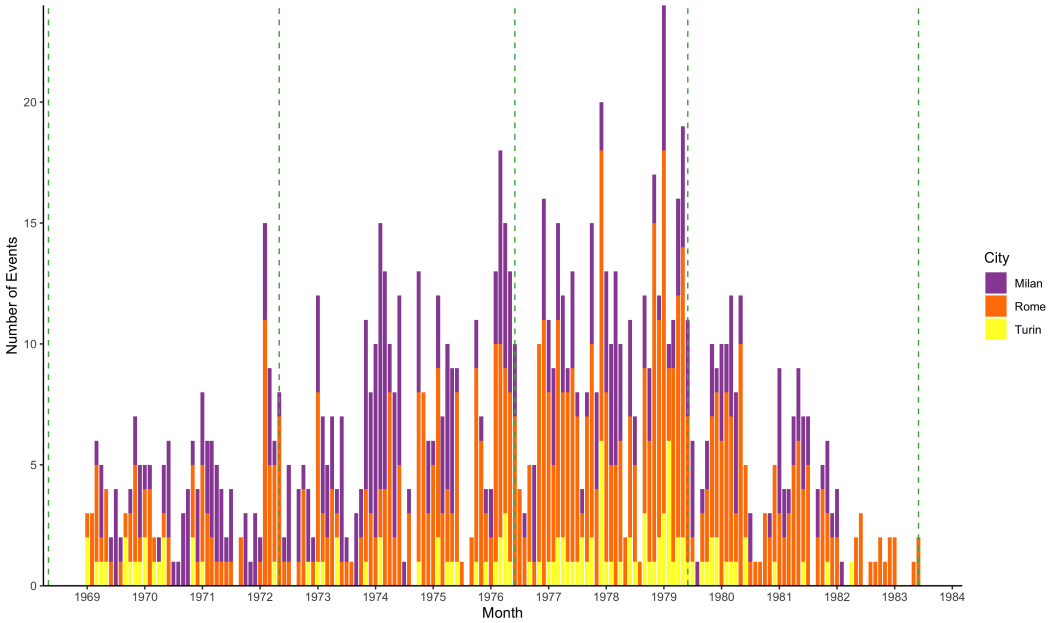
C. Appendix: Figures

Figure A2: Number of terrorist events by political faction



Notes: The bar plot displays the total number of terrorist events categorized according to the political faction of the perpetrators. Data have been digitalized by the author from ISODARCO (1992).

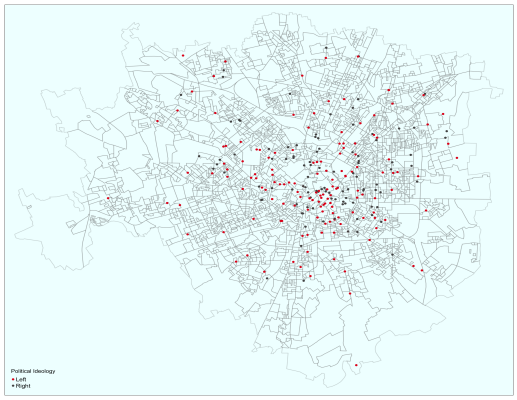
Figure A3: Number of terrorist events by City



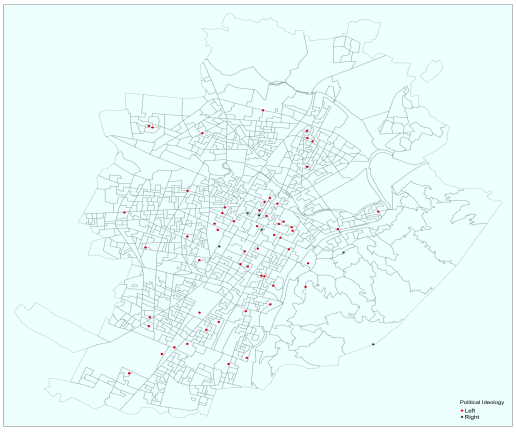
Notes: The bar plot displays the total number of terrorist events categorized according to the political faction of the perpetrators. Data have been digitalized by the author from ISODARCO (1992).

Figure A4: Spatial distribution of events

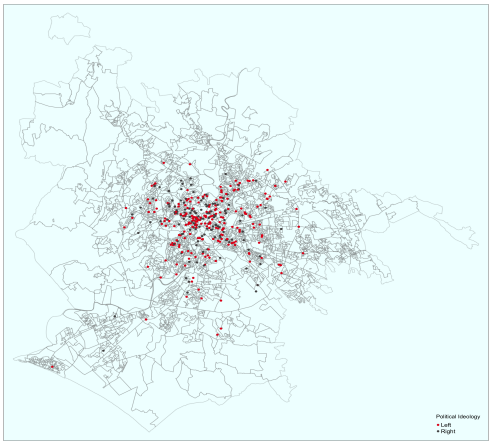
(A) Milan



(B) Turin



(C) Rome



Notes: The maps display the spatial distribution of the terrorist attacks split for the three municipalities in the sample. The underlined shapefiles are the boundaries of electoral precincts as constructed by Pinto (2023). The events have been taken from ISODARCO (1992) and geocoded by the author using the Google Maps API.

D. Appendix: Tables

In this appendix, additional tables are reported to test the robustness of the main effects discussed in the paper. Table A2 presents the impact of smaller-scale violence committed by right-wing and left-wing extremists who are not affiliated with more structured terrorist organizations. Tables A3 and A4, report robustness for the main estimates changing the radius used in the baseline. Tables A5 and A6, report robustness for the main estimates changing the time window for treatment to two years and one year before the election date respectively. Overall results are consistent with the estimates reported in the main body of the paper.

Table A2: Political violence and ideological roots: impact on parties votes

	Δ DC-PCI	DC	PCI	Δ DC-PCI	DC	PCI
	(1)	(2)	(3)	(4)	(5)	(6)
# Attacks Right	0.0020*	0.0016*	-0.0004			
	(0.0011)	(0.0009)	(0.0005)			
# Attacks Left				0.0047***	0.0047***	0.0001
				(0.0015)	(0.0011)	(0.0007)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Municipality-Election	✓	✓	✓	✓	✓	✓
Dep. Var. Mean	-0.0012	0.2909	0.2921	-0.0012	0.2909	0.2921
R^2	0.8747	0.7424	0.8965	0.8748	0.7427	0.8965
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (4) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (5) the dependent variable is the votes' share for the Christian Democracy, while in columns (3) and (6) the dependent variable is the votes' share for Communist Party respectively. # Attacks Right is the number of political violence attacks perpetrated by far-right militants not organized in proper terrorist groups, happening before an election e within 500 meters from a given precinct. # Attacks Left is the number of political violence attacks perpetrated by far-left militants not organized in proper terrorist groups, happening before an election e within 500 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.1 Robustness Radius and Time Window

Table A3: Political violence impact on party votes: alternative radius 600m

	Δ DC-PCI	DC	PCI	Δ DC-PCI	DC	PCI
	(1)	(2)	(3)	(4)	(5)	(6)
I(attack)	0.0028**	0.0023***	-0.0005			
	(0.0013)	(0.0009)	(0.0008)			
# Attacks				0.0021***	0.0021***	-0.0001
				(0.0005)	(0.0004)	(0.0002)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Election-Municipality	✓	✓	✓	✓	✓	✓
R^2	0.87474	0.74241	0.89651	0.87485	0.74295	0.89651
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (4) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (5) the dependent variable is the votes' share for the Christian Democracy, while in columns (3) and (6) it is the vote share for the Communist Party. *I(attack)* is a binary variable taking value 1 if in the period prior to an election e at least one attack happens within 600 meters from a given precinct and 0 otherwise. *# Attacks* is the number of attacks happening before an election e within 600 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Political violence impact on party votes: alternative radius 400m

	Δ DC-PCI	DC	PCI	Δ DC-PCI	DC	PCI
	(1)	(2)	(3)	(4)	(5)	(6)
I(attack)	0.0021	0.0025**	0.0004			
	(0.0014)	(0.0010)	(0.0008)			
# Attacks				0.0023***	0.0024***	0.0001
				(0.0007)	(0.0005)	(0.0004)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Election-Municipality	✓	✓	✓	✓	✓	✓
R^2	0.87472	0.74241	0.89651	0.87477	0.74265	0.89651
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (4) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (5) the dependent variable is the votes' share for the Christian Democracy, while in columns (3) and (6) it is the vote share for the Communist Party. *I(attack)* is a binary variable taking value 1 if in the period prior to an election e at least one attack happens within 400 meters from a given precinct and 0 otherwise. *# Attacks* is the number of attacks happening before an election e within 400 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Political violence impact on party votes - 2 years time window

	Δ DC-PCI	DC	PCI	Δ DC-PCI	DC	PCI
	(1)	(2)	(3)	(4)	(5)	(6)
I(attack)	0.0049***	0.0033***	-0.0017**			
	(0.0014)	(0.0010)	(0.0008)			
# Attacks				0.0024***	0.0023***	-0.0001
				(0.0007)	(0.0005)	(0.0004)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Election-Municipality	✓	✓	✓	✓	✓	✓
R^2	0.87477	0.74246	0.89652	0.87477	0.74257	0.89651
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (4) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (5) the dependent variable is the votes' share for the Christian Democracy, while in columns (3) and (6) it is the vote share for the Communist Party. *I(attack)* is a binary variable taking value 1 if in the two years prior to an election e at least one attack happens within 500 meters from a given precinct and 0 otherwise. *# Attacks* is the number of attacks happening in the two years before an election e within 500 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Political violence impact on party votes - 1 year time window

	Δ DC-PCI	DC	PCI	Δ DC-PCI	DC	PCI
	(1)	(2)	(3)	(4)	(5)	(6)
I(attack)	0.0048***	0.0045***	-0.0003			
	(0.0017)	(0.0013)	(0.0009)			
# Attacks				0.0022*	0.0029***	0.0007
				(0.0012)	(0.0009)	(0.0006)
Fixed Effects:						
Precinct	✓	✓	✓	✓	✓	✓
Election-Municipality	✓	✓	✓	✓	✓	✓
R^2	0.87475	0.74250	0.89651	0.87473	0.74247	0.89651
Precincts (clusters)	6,182	6,182	6,182	6,182	6,182	6,182
Observations	29,661	29,661	29,661	29,661	29,661	29,661

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a precinct and election. In columns (1) and (4) the dependent variable is the difference in vote share between the Christian Democracy and the Communist Party. In columns (2) and (5) the dependent variable is the votes' share for the Christian Democracy, while in columns (3) and (6) it is the vote share for the Communist Party. *I(attack)* is a binary variable taking value 1 if in the year prior to an election e at least one attack happens within 500 meters from a given precinct and 0 otherwise. *# Attacks* is the number of attacks happening during the year before an election e within 500 meters from a given precinct. In all columns precinct and municipality-election fixed effects are included. The sample covers the five parliamentary elections for the Italian lower chamber (Camera dei Deputati) held over the years 1968-1983. Electoral data have been digitized by the author from registries provided by the Historical Archive of the Deputy Chamber as detailed in Appendix B. The number of attacks has been obtained from ISODARCO (1992) as detailed in Appendix B. Standard errors are clustered at the precinct level. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

CHAPTER 3

Water Wars

Devis Decet
Northwestern University

Andrea Marcucci
University of Lausanne

Abstract

We study the relationship between access to water resources and local violence in Africa. Due to limited irrigation, rural communities rely on rainfall, rivers, and lakes for their economic needs. Rainfall scarcity can make access to water from rivers and lakes more valuable, thereby generating conflicts in rural settings. We explore this hypothesis by integrating granular data on the river network with high-resolution data on rainfall and violent conflict events in Africa from 1997 to 2021. We find that reduced rainfall in a location leads to more conflict in neighboring areas that are water-rich and located upstream along the river network. These are the sites that exert more control over the river flow. The effect is more pronounced in regions experiencing a long-term decline in water presence. Consistent with the proposed mechanism, conflicts concentrate in areas with higher returns to water access, as proxied by the presence of agricultural production. Additionally, the impact is more pronounced in regions with unequal water distribution among ethnic groups, highlighting how cooperation costs are an important friction preventing peaceful sharing of water resources. In terms of policy responses, we find that the effects tend to be mitigated in countries with stronger democratic institutions, better rule of law, higher state capacity and less corruption.

Keywords: Conflict, water, climate change, rivers, resource competition, Africa.

JEL: D74, Q25, N47, O13, Q34.

1. Introduction

Access to water is essential for human life and economic activity. Estimates suggest that four billion people experience at least one month per year without access to sufficient water (Mekonnen and Hoekstra, 2016). Climate change is likely to exacerbate this situation, thereby drawing attention to the potential for conflicts over access to water (United Nations, 2023; World Economic Forum, 2023). This is a natural concern, given the role of climatic shocks and competition over resources in fostering violence (Burke, Hsiang and Miguel, 2015; McGuirk and Burke, 2020). However, we lack systematic evidence on whether, and how, climatic shocks can induce conflicts over water resources.

In rural Africa, where the economy is largely dependent on agriculture and pastoralism, this issue is particularly salient. Due to the lack of large irrigation infrastructures, these economic activities rely mainly on rainfall, wells and surface water. In this context, those residing close to rivers and lakes can use surface water for their needs. For instance, farmers construct irrigation channels from rivers or practice recession agriculture, which involves cultivating lands enriched by river sediments. Pastoralists similarly exploit these water bodies as drinking points for their livestock.

This paper investigates systematically the occurrence of conflicts over water resources in Africa from 1997 to 2021. There are specific locations and time periods where we expect conflicts over water resources to occur. They are more likely to arise during years of low rainfall, when the value of accessing surface water increases. In such scenarios, drought-affected individuals are likely to seek water access in adjacent, water-abundant cells. Additionally, those experiencing a drought primarily contend for access to water in upstream locations, as upstream they can exert more control over the river flow and water quality is generally better. Summing up this argument, we expect that a location is more likely to experience conflict over water resources if it is water rich and a drought happens in a region located downstream.

In our empirical analysis, we bring this argument to the data. Utilizing cells of $0.5^\circ \times 0.5^\circ$ degrees in latitude and longitude as units of observation, we measure the incidence of conflict using geocoded event data across all African countries from Armed Conflict Location Events Data Project (ACLED), which provides details on the date, location, and type of conflicts. For assessing surface water resource distribution, we employ hydrological data from the Global Floods Awareness System (Harrigan, Zsoter, Alfieri, Prudhomme, Salamon, Wetterhall, Barnard, Cloke and Pappenberger, 2020). Additionally, we rely on the HydroBASINS dataset (Döll, Kaspar and Lehner, 2003) to determine for each pair of cells their up-downstream relationship along the rivers network.

For each cell, we define its neighborhood as all surrounding cells within a 180 km radius and assign a measure of water richness to the cell itself. Our preferred measure is *Water Discharge*,

representing the annual average water flow through a cell. We assess whether the impact on violence of low rainfall in a downstream neighboring cell is amplified in cells that are *water rich*. By employing geographically disaggregated data, we can estimate a specification that includes grid-cell fixed effects, to account for local time-invariant factors, and country-year fixed effects, to control for common macro-level factors that vary by country and year. Our approach also allows us to control for any direct effects of rainfall occurring in the grid-cell itself.

Our main result is that rainfall shocks in a downstream cell increase the likelihood of conflict differentially more in locations that have higher *Water Discharge*. Our preferred specification implies that when a downstream cell experiences a rainfall shock, the likelihood of conflict is 0.6 percentage points larger for a cell with high *Water Discharge*, compared to one with low *Water Discharge*, corresponding to 7.30% of the conflict incidence mean.[1] These findings are robust to alternative coding of the water richness measure, to using alternative conflicts data and to controlling for other relevant confounders like temperature and population.

In light of these findings, we further delve into the economic incentives behind conflicts over water resources. We expect a higher likelihood of conflict in areas where the benefits of water access are larger. Given Africa's predominantly agrarian economy, water exploitation is primarily linked to agricultural activities. Therefore, cells with significant agricultural output are likely to offer higher returns from water access. To investigate this channel we split the sample between cells with high and low agricultural production. We find that, consistently with our expectations, the effects are driven by places characterized by higher presence of agriculture.

Surface water resources may be distributed unequally across space, yet individuals from different areas can cooperate and manage them together. We thus expect that conflict arises in contexts in which the costs of cooperation are higher. To explore this possibility, we use data from Giuliano and Nunn (2018) to identify the linguistic groups residing in each cell. Ethnic grievances might imply too high cooperation costs. Indeed, we observe that effects are stronger in areas with more unequal distribution of water resources across different ethnic groups, measured as polarization, Gini and Theil indexes. This evidence suggests that high cooperation costs contribute to the failure of peaceful water sharing and lead to an increased reliance on violence.

Climate change poses a multifaceted threat to water resources, not only through more frequent droughts but also via long-term desertification processes. In order to examine this extended impact, we have developed a simple metric for long-term water depletion. We find

[1] For this quantification high discharge corresponds to the third quartile of the discharge distribution, whereas low discharge corresponds to the first quartile level.

that the effect of droughts on conflict over water resources is stronger in regions where water availability has diminished over the past forty years. The findings suggest that adaptation costs exacerbate the problem, and that conflict over water resources may become a more urgent issue as climate change intensifies the desertification process in certain areas.

In the final section of the paper, we explore the ability of formal institutions to mitigate conflicts induced by climate change. Stronger state presence might be essential not only for the effective redistribution of resources but also for the development and implementation of infrastructural solutions designed to mitigate crisis situations. We look at whether countries with better institutional characteristics are less likely to experience conflict over water resources. Considering various measures of formal institutions, such as democratic governance, rule of law, absence of corruption and government effectiveness, reveals a consistent pattern: conflicts triggered by droughts are primarily a concern in countries with relatively weaker institutions.

Our research contributes to the literature on climate and conflict by presenting new evidence that identifies a precise mechanism through which climate change (Hsiang and Burke, 2014; Burke, Hsiang and Miguel, 2015) and weather shocks (Miguel, Satyanath and Sergenti, 2004; Sarsons, 2015; Almer, Laurent-Lucchetti and Oechslin, 2017; Unfried, Kis-Katos and Poser, 2022) influence local violence. Recent works by Eberle, Rohner and Thoenig (2020) and McGuirk and Nunn (2020) have emphasized the impact of heat and changing rainfall patterns on conflicts between farmers and pastoralists. In our study, we focus on the effects of low rainfall years, which are becoming more frequent in Africa due to climate change, and how they increase competition for accessing and controlling surface water resources. A key aspect of our analysis involves investigating spillovers, wherein low rainfall in one area leads to heightened conflict in water-rich territories located upstream. By identifying this specific mechanism, we provide insights into the spatial spillovers observed in existing climate-conflict research (Guariso and Rogall, 2017; Harari and Ferrara, 2018). In doing so, we complement the specific mechanisms of conflicts diffusion studied by König, Rohner, Thoenig and Zilibotti (2017) and McGuirk and Nunn (2020).

We also speak to the literature on the determinants of conflict, which has focused on the importance of ethnic or social factors (Esteban, Mayoral and Ray, 2012; Rohner, Thoenig and Zilibotti, 2013; Depetris-Chauvin and Özak, 2020; Moscona, Nunn and Robinson, 2020; Arbatlı, Ashraf, Galor and Klemp, 2020), of historical factors (Besley and Reynal-Querol, 2014; Michalopoulos and Papaioannou, 2016; Depetris-Chauvin, 2015), and economic factors, especially shocks to resources value and conflict opportunity cost (Dube and Vargas, 2013; Berman, Couttenier, Rohner and Thoenig, 2017; McGuirk and Burke, 2020; Adhvaryu, Fenske, Khanna and Nyshadham, 2021). We demonstrate that controlling and accessing surface water resources can be a determinant of conflict.

In a nutshell, our paper contribution to the literature of the economics of conflict is manifold. To the best of our knowledge, we are the first to show that the control of surface water resources is a mechanism linking climate shocks and conflict. Additionally, we find that, under unfavorable climatic conditions, water can induce a resource curse. Finally, leveraging on new fine grained data, we document how the rivers network structure can shape the spatial spillovers observed in existing climate-conflict research.

The remainder of the paper is organized as follows. Section 2 provides a description of the context and of how rivers and lakes' water is used for economic activity in rural Africa. In Section 3 we introduce our data sources and we detail how we build the variables used in the analysis. Section 4 describes the empirical strategy and the results of the paper. Finally, Section 5 concludes.

2. Background and Context

2.1 Using surface water resources to smooth water consumption

Water is an essential resource for agriculture, pastoralism, and daily consumption. In rural Africa, the absence of infrastructures such as piped water and irrigation systems necessitates heavy reliance on rainfall, wells, and surface water. In this context, we provide examples illustrating how households utilize surface water resources for their economic activities and everyday life. Our aim here is to illustrate concretely the significance of controlling water resources.

An example is flood-based farming systems (for more details refer to Puertas, van Steenbergen, Haile, Kool and Embaye, 2021). This agricultural practice capitalizes on the nutrient-rich soil deposited by river floods. Another variant of this approach is the use of inundation canals, where land is irrigated through canals supplied by temporary high water levels in rivers. These methods become particularly crucial during low rainfall periods, stressing the importance for farmers to maintain control over land near surface water sources, enabling them to effectively utilize these agricultural techniques.

In general, the construction of canals plays a vital role in bringing water from rivers to arid regions. An example of this is the initiative undertaken by the World Food Programme (WFP) in Kenya, where paved canals have been built from the Turkwell River. These canals efficiently channel water to farms in neighboring areas, benefiting over 45,000 farmers. As a result, farmers can effectively irrigate their fields even during seasons with limited rainfall (World Food Programme, 2023). Farmers located near rivers have the advantage of lower canal construction costs and can harness the water flowing through them to a greater extent.

Likewise, water resources are crucial for pastoralists. Rivers and lakes act as natural hydration points for livestock, and the areas around these water bodies often maintain vegetation even

in dry seasons. This availability of vegetation enables herders to provide reliable nourishment for their livestock.

Securing land along a river grants farmers enhanced access to water resources, yet such control can significantly affect water availability further downstream. One extreme example is the Omo River which flows between Ethiopia and Kenya (Climate Diplomacy, 2023*c*). In the rural communities of the Lower Omo River Valley, a combination of flood recession agriculture and pastoralism is practiced, both of which depend on the seasonal floods of the Omo River to replenish crop and grazing lands along the riverbank. The establishment of irrigated sugar plantations in Ethiopia (situated upstream) has the potential to impact the water availability in these regions, as water diversion for these plantations can disrupt the natural flow downstream.

2.2 Climate change and conflicts over water resources

Freshwater resources may be distributed unequally, yet different groups can cooperate and manage them together. For instance, according to the hydraulic theory, the formation of early states was partly motivated by the necessity of institutions for large-scale irrigation projects.[2] Moreover, a symbiotic system has often existed between farmers and herders, with herders migrating to farmers' land during dry seasons. This traditional arrangement, especially when farmers' land is situated near rivers, can be seen as a norm that enables efficient sharing of water resources among different groups during periods of limited rainfall.

However, climate change-induced rainfall scarcity in Africa is undermining these established water-sharing institutions, leading to their deterioration. For instance, herders migrate earlier to water-rich lands, causing conflicts with farmers still cultivating crops (Eberle et al., 2020 and McGuirk and Nunn, 2021). Additionally, farmers may extract more water for irrigation during rainfall shortages, reducing downstream water flow. As recently happened in Laikipia county, in Kenya, or in Fayoum, in north Egypt, this can induce groups located downstream to resort to violence to destroy the irrigation infrastracture or scare the farmers upstream, especially if the government does not take actions (Nation, 2023, Monitor, 2022). Climate change also creates new situations requiring cooperation over water resources without pre-existing arrangements. A notable example of this is observed when droughts force pastoral groups to modify their migratory routes, often leading to competition with other pastoralists over the same water sources. An illustration of this situation can be found in the Lower Omo and Turkana region along the Kenyan-Ethiopian border. Local communities in search of water and grazing land have expanded their ranges, leading to increased proximity and frequent

[2] See Allen, Bertazzini and Heldring (2020) for econometric evidence supporting this theory in the case of ancient Mesopotamia.

clashes with other groups over these resources. From 1989 to 2011, conflicts between the Nyangatom, Daasanach, and Turkana groups alone resulted in over 600 direct deaths (Climate Diplomacy, 2023*b*).

3. Data

This section describes the data sources and the construction of the variables used in the analysis. Our empirical analysis is based on a geo-referenced, annual panel that divides the African continent into 10,229 grid cells (see Figure A2). These grid cells have a size of $0.5° \times 0.5°$ degrees, equivalent to approximately 55 km $\times$ 55 km at the equator. Throughout our analysis, the unit of observation is a cell-year pair.

3.1 Data Sources

Conflict

Our study utilizes georeferenced conflict events from the Armed Conflict Location & Event Data Project (ACLED) covering the period from 1997 to 2021 (Raleigh, Linke, Hegre and Karlsen, 2010). The ACLED data has no requirement for a specific number of fatalities within a calendar year or for a conflict event. As a result, the ACLED data is very apt for capturing smaller-scale, localized conflict events. ACLED gathers information on conflict events from multiple sources, including regional and national media outlets, NGOs, and humanitarian organizations. The ACLED data includes the date and geographic coordinates of each event. We retain only events that are precisely geolocalized. In our main analysis, we consider only events categorized as "battles", and "violence against civilians", excluding thereby less violent events like "riots" and "protests". In fact, according to the mechanism we are considering when a shock occurs, individuals tend to move upstream to access water resources, resulting in the emergence of more lethal and intense conflicts compared to mere riots or protests.[3] Figure A3 reports the average yearly incidence for ACLED conflict data.

In some robustness checks, we use georeferenced conflict events from the Uppsala Conflict Data Program (UCDP) (Sundberg and Melander, 2013) covering the period from 1989 to 2020. In the UCDP data, conflict events are characterized as either two-sided battles or one-sided attacks that fulfill specific criteria. In order to be included, a conflict event must involve at least one fatality, and the conflict dyad (i.e., the pair of actors involved) must have caused a minimum of 25 fatalities within at least one calendar year during the series. Moreover, at least one of the actors involved must be an "organized actor," such as a state or a politically organized rebel group or militia. These data are compiled following a two-step process, by

[3] In Table A4 we show that we do not observe any effect using riots or protests incidence as dependent variables.

which global newswire sources are consulted first, and then confirmed consulting local/specialized sources, such as translations of local news performed by the BBC, local media, NGO reports, and field reports. Like ACLED data, UCDP data includes the date and geographic coordinates of each event. We consider only precisely geolocalized events.[4]
By utilizing the date and geographic location (longitude and latitude) we are able to assign each event to a specific cell-year pair. For both data sources, we aggregate the information at the cell-year level. We code conflict incidence as 1 if any conflict event occurred within a cell-year and as 0 otherwise.

Hydrology

In our analysis, we include data on river discharge obtained from the Global Floods Awareness System.[5] River discharge refers to the volume of water passing through the section of a river per unit of time, measured in cubic meters per second. The data we utilize provides daily average river discharge on a global scale, with a spatial resolution of $0.05° \times 0.05°$ decimal degrees. The data are produced by combining information from satellites, in-situ measurements, and hydrological models. Notice that the quantity of water reported in the data takes into account all types of surface water bodies, including lakes, ponds, rivers and streams. We aggregate this information at the cell-year level (see Figure A4). To incorporate information on the river network topology, we rely on the HydroBASINS dataset.[6] This dataset offers a shapefile of drainage basins, which are globally consistent geospatial units frequently employed in environmental and hydrology studies. Each basin represents the land area that collects and channels precipitation, such as a valley. The shapefiles are available at different levels of aggregation; we use level 7 as in Eberle (2020). We allocate each cell to a specific basin based on the amount of water in the overlapping area. Specifically, for every intersection between a river basin and a square grid cell, we assign the cell to the basin if that particular intersection contains the greatest amount of water. Then, we construct a matrix that describes the relationship between each pair of cells along the rivers' network exploiting the Pfafstetter coding system.[7] This matrix enables us to identify whether a pair of cells is connected upstream, downstream, or not connected at all.[8] We are the first, to the best of our knowledge, to use the Pfafstetter coding system to pin down the upstream-downstream

[4]To be more specific, our analysis includes only those events that have been geolocated with a minimum precision of the town level (precision level 3).

[5]Accessible from Harrigan et al. (2020).

[6]Part of the HydroSHEDS environment and accessible from `https://www.hydrosheds.org/`; for further details, see Döll et al. (2003).

[7]The Pfafstetter coding system is widely used in hydrology to determine the up-downstream relationship between rivers' basins, see for example Verdin and Verdin (1999).

[8]See appendix Section A for further details about the construction of the river network matrix.

relationships between uniform squared cells. This allows us to employ the standard units of observations from the conflict literature and at the same time to integrate them with the spatial structure imposed by the rivers network.

Rainfall

Following Harari and Ferrara (2018) we use precipitation data from ERA5 (Hersbach, Bell, Berrisford, Biavati, Horányi, Muñoz Sabater, Nicolas, Peubey, Radu, Rozum, Schepers, Simmons, Soci, Dee and Thépaut, 2023). ERA5, a reanalysis dataset, offers comprehensive weather data for the period 1959 through 2021. It provides data at various grid resolutions and temporal resolutions as fine as 6 hours. The dataset is derived from a combination of high-frequency observations collected from diverse sources, including weather stations, satellites, and probes. ERA5 represents a notable improvement over gauge data, particularly in regions with limited weather station coverage like Africa. In fact, it is important for us not to rely exclusively on raw gauge data for two reasons. Firstly, due to the scarcity of weather stations across Africa, extensive interpolation would be required, potentially resulting in artificial patterns of spatial correlation in weather shocks. Secondly, the availability of gauge data itself may be influenced by the presence of conflict.

Other Data

We assign ethnic groups to territories across the continent using the geographic distribution of linguistic groups from Giuliano and Nunn (2018). These data are built by linking manually ethnic groups to languages and dialects; the geographic distribution of languages and dialects is from Gordon and Grimes (2009). Additionally, we use information about agricultural land cover from the replication package of McGuirk and Burke (2020). Temperature data are from Hersbach et al. (2023). Population data are from Center for International Earth Science Information Network - CIESIN - Columbia University (2018). Finally, state capacity indicators are taken from Kaufmann, Kraay and Mastruzzi, 2011.

3.2 Variables Definition

For each cell of our grid, we define a neighborhood as all the cells in a 180 km radius. We choose this buffer to be consistent with the seminal work of Harari and Ferrara (2018). In the top panel of Figure A1 we report an example of how we build a neighborhood (all the highlighted cells) around the reference cell (in dark yellow). Then, as a way of example, we overlap hydrographic data of a section of the Niger River with our grid, showing how we establish which cells are upstream (orange) or downstream (red) to the reference cell (see appendix Section A for further details).

Water Richness

We propose different definitions of water richness. Notice that these measures change over time, because of the time-varying dimension of our hydrological data. Our preferred measure of water presence is *Water Discharge*, a continuous measure of water abundance corresponding to the natural logarithm of the mean amount of freshwater present in a cell during a year (see Figure A4). More precisely, it is the sum of the water passing through the sections of all the rivers flowing in a given cell, measured in cubic meters per second. In order to understand the impact of shocks in places that are extremely water rich, we also consider two alternative measures using a simple dummy. *Water Monopolist* (see Figure A5) is a dummy which indicates cells that have the largest quantity of water in their neighborhood. To be specific, *Water Monopolist* is equal to one for cell i in neighborhood n, if *Water Discharge* of cell i is the highest of the neighborhood. Finally, *Water Monopolist* + (see Figure A6) requires the additional condition that a cell has abundance of water also in absolute terms. Specifically, *Water Monopolist* + takes value one for cells which are *Water Monopolist* and whose *Water Discharge* is above the median of the continent in a year.

Rainfall Shocks

To identify rainfall shocks, we adopt the methodology outlined in Burke, Gong and Jones (2015) and Corno, Hildebrandt and Voena (2020). We utilize a long-term time series spanning from 1959 to 2021, consisting of rainfall observations. For each geographical cell, we fit a gamma distribution to the calendar year rainfall data. This distribution estimation allows us to characterize the typical rainfall patterns for a specific location. Using the estimated gamma distribution, we determine which location-years experienced rainfall levels below the 15th percentile of the distribution. We code these instances as rainfall shocks.

Ethnic Inequality in Water Access

An imbalanced allocation of water among distinct ethnic groups could potentially hinder the sharing of this resource in case of adverse climate shocks raising cooperation costs. To gain a better understanding of this process, we calculate various indexes that describe water allocation within each neighborhood among different ethnic groups. Specifically, we use the geographic distribution of linguistic groups from Giuliano and Nunn (2018) and overlap it with the shapefile of each cell's neighborhood. For each linguistic group within a neighborhood, we determine their water ownership. Finally, we calculate different statistics at the neighborhood level based on the water ownership of each ethnic group. In particular, we compute polarization following the index proposed by Reynal-Querol (2002). Differently from the original measure, which relies on population shares, our approach considers water shares as a proportion of the total water quantity in a neighborhood. Consequently, the index takes its

maximum value if in a given area there are only two groups owning 50% of the total water amount. As alternative measures to account for inequality in water distribution between ethnic groups we compute the Gini and Theil indexes. We report the spatial distribution of these three variables in Figure A7.

4. Rainfall Scarcity and conflict over water resources

4.1 Empirical Strategy

Our objective is to test systematically the occurrence of conflicts related to water resources at a local level. There are specific locations and time periods where we expect conflicts over water resources to occur. They are more likely to arise during years of low rainfall, when the value of surface water increases. It is in such cases that individuals affected by drought conditions are more inclined to seek access to water in neighboring cells, particularly if these cells are abundant in water.

Additionally, those experiencing a drought primarily contend for access to water in upstream locations, as upstream they can exert more control over the river flow and water is normally more abundant and of higher quality. Summing up this argument, we expect that a cell is more likely to experience conflict over water resources if it is water rich and a drought happens in a cell located downstream.

We present here our baseline equation which estimates whether adverse rainfall shocks in downstream territories have a differentially higher impact on cells that are water rich.

$$y_{it} = \lambda_1 \text{Water Rich}_{it} + \lambda_2 Shock_{it}^{Down} + \beta Shock_{it}^{Down} \times \text{Water Rich}_{it} + \mathbf{X}'_{it}\Gamma + \mu_i + \mu_{ct} + \varepsilon_{it} \quad (1)$$

Where y_{it} is a dummy variable for conflict incidence in cell i during year t, Water Rich$_{it}$ is a time varying measure of water richness in a given cell, and $Shock_{it}^{Down}$ takes value one if a cell in the neighborhood located downstream to cell i is hit by a rainfall shock during year t. We include in the regression cell fixed effects μ_i and country-year fixed effects μ_{ct} to account for time invariant cell characteristics and country specific yearly shocks that might affect conflicts. $\mathbf{X}'_{it}$ are additional cell specific time-varying variables. In some specifications, we control for rainfall shocks happening in cell i, which may have direct effects on local violence. We also show that results are unaffected by including rainfall shocks happening in cells located upstream to i, and we allow them to have differential impact depending on water presence (Water Rich$_{it}$). In sensitivity analysis we include additional time varying controls, that we introduce in Section 4.3.

Our hypothesis is that if a drought happens downstream, water rich cells are more likely to

experience conflict. Thus, we expect $\beta > 0$.

4.2 Baseline Results

In Table 1 we present results with our preferred measure of water richness: *Water Discharge*. *Water Discharge* corresponds to the average quantity of water present in a cell during a given year.

In column 1 we estimate the main regression equation 1, testing our hypothesis that a cell is more likely to experience conflict over water resources if it is water rich and a drought happens in a cell located downstream. The coefficient β is positive and statistically significant at the 1% level. In column 2 we check whether our hypothesis that only downstream shocks have an impact on conflict incidence is valid interacting our measure of water presence with shocks happening upstream. We cannot find any significant impact of upstream shocks on conflict. In columns 3 and 4 we control for any direct effect of rainfall shock happening in the cell, results are unaffected. Finally, in column 5 we test whether groups located downstream and upstream have different incentives to fight when hit by a drought, including both shocks in the same regression. We can appreciate how upstream shocks do not display the same patterns as downstream shocks. Moreover, our coefficient of interest is very stable and, if anything, it becomes larger and more precisely estimated. Interpreting the magnitude of the coefficient in our preferred specification (column 5) we have that when a downstream cell experiences a rainfall shock, the likelihood of conflict is 0.6 percentage points higher for a cell with high *Water Discharge*, compared to one with low *Water Discharge*,[9] corresponding to 7.30% of the dependent variable mean. This is in line with a predatory mechanism of seeking control over the water flow of the river when the resource becomes scarcer. In Table A2, we include the same specifications of Table 1 but reporting spatially clustered standard errors, allowing for a spatial correlation within a 500 km radius of a cell's centroid and infinite serial correlation (Conley, 1999). While the estimates become generally less precise, the coefficient β from equation 1 retains statistical significance at the 5% level.

4.3 Sensitivity Analysis

Alternative Measures of Water Richness

In Table 2 we estimate equation 1 using all the three measures of water presence. Column 1 corresponds to column 5 of Table 1, while in columns 2 and 3 we interact weather shocks with a binary variable. Specifically, in column 2, *Water Measure* takes value one if a cell is the one

[9] For this quantification high discharge corresponds to the third quartile of the discharge distribution, whereas low discharge corresponds to the first quartile level.

Table 1: Precipitation shocks and water discharge

	Incidence (ACLED)				
	(1)	(2)	(3)	(4)	(5)
Water Discharge	0.0010	0.0007	0.0010	0.0009	0.0009
	(0.0009)	(0.0010)	(0.0009)	(0.0009)	(0.0009)
Water Discharge × Shock Down	0.0011***		0.0011***		0.0012***
	(0.0004)		(0.0004)		(0.0004)
Water Discharge × Shock Up		0.0003		0.0003	-0.0002
		(0.0005)		(0.0005)	(0.0005)
Shock Down	0.0008		0.0010		0.0009
	(0.0017)		(0.0018)		(0.0018)
Shock Up		-0.0018		-0.0024	-0.0014
		(0.0020)		(0.0021)	(0.0021)
Shock Own			-0.0005	0.0020	0.0000
			(0.0017)	(0.0016)	(0.0017)
Cell FE	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201	0.08201	0.08201
R^2	0.42101	0.42095	0.42101	0.42096	0.42101
Cells	10,228	10,228	10,228	10,228	10,228
Observations	255,700	255,700	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in Section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

with the highest water discharge in its neighborhood during a given year. Finally, in column 3, we focus on cells that not only have the highest water discharge in their neighborhood but also exceed the median value in the sample. This methodology effectively excludes cells with minimal discharge, in particularly desert areas. In all columns, we observe a positive coefficient for the interaction term between downstream precipitation shocks and water presence. In cells with particularly high levels of water presence (column 3) a precipitation shock causes an increase in conflicts of 3.4 percentage points, which corresponds to about 42% of the dependent variable mean. Reassuringly, all the three measures aimed at capturing water richness deliver consistent results.

Table 2: Precipitation shocks all measures

	Incidence (ACLED)		
	Water Discharge	Water Monopolist	Water Monopolist +
	(1)	(2)	(3)
Water Measure	0.0009	0.0120	0.0151
	(0.0009)	(0.0098)	(0.0106)
Water Measure × Shock Down	0.0012***	0.0181	0.0336**
	(0.0004)	(0.0123)	(0.0170)
Water Measure × Shock Up	-0.0002	-0.0020	-0.0046
	(0.0005)	(0.0118)	(0.0144)
Shock Own	0.0000	-0.0004	-0.0004
	(0.0017)	(0.0017)	(0.0017)
Shock Down	0.0009	0.0049***	0.0048***
	(0.0018)	(0.0015)	(0.0015)
Shock Up	-0.0014	-0.0018	-0.0018
	(0.0021)	(0.0017)	(0.0017)
Cell FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201
R^2	0.42101	0.42101	0.42103
Cells	10,228	10,228	10,228
Observations	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. *Water Measure* indicates generically a measure of water quantity which varies between columns. In column (1) it is the natural logarithm of the average water discharge present in a cell during a given year (*Water Discharge*). In column (2) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year (*Water Monopolist*). In column (3) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year and the discharge is higher than the median level in the sample for that year (*Water Monopolist +*). *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in Section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Alternative conflict dataset

In Table A3 we replicate our baseline analysis using alternative conflict data from the UCDP georeferenced Event Dataset (Sundberg and Melander, 2013) that focuses on violence perpe-

trated by larger-scale and more structured groups. Our coefficient of interest remains positive, large and precisely estimated in all three specifications.

Alternative conflict definitions

In Table A4 we replicate our main regression results using different conflict categories from the ACLED dataset. In column 1 we replicate column 5 of Table 1, in column 2 we consider only battles (the most deadly type of conflicts present in our data) in column 3 other kind of violent attacks against civilians by organized groups, while in the last two columns we look at less intense and deadly conflict types like protests (column 4) and riots (column 5). In line with the mechanism we have in mind, we observe an effect only for larger scale type of conflicts. Individuals do not move upstream just for rioting or protesting against the government, but to fight over access to water resources.

Additional controls

In Table A9 we show that our results are robust to controlling for other factors which have been associated with conflict. Specifically, we control for (log) population, yearly average temperature and yearly average temperature during the day. Finally, we check whether results are robust to controlling for lagged conflict incidence. The estimates of our main coefficient of interest are unaffected by the inclusion of the controls.

Alternative neighborhood and rainfall shocks

In the appendix, from Table A5 to Table A8, we conduct additional robustness checks to ensure that our results are not influenced by the specific parameter choices we have made. Specifically, Tables A5 and A6 explore alternative thresholds for defining a rainfall shock, using different percentiles as cutoff points in the distribution. Conversely, Tables A7 and A8 examine the effects of using alternate radii of 160 km and 200 km, respectively, to define a cell neighborhood. Across all these analyses, our primary coefficient of interest maintains a magnitude and significance level similar to that estimated in our main specification (column 5 of Table 1).

4.4 Heterogeneous characteristics affecting conflicts

In this section, we explore the characteristics that increase the likelihood of conflicts arising over water resources.

Agricultural land and returns to water access

Table 3: Agricultural Land

	Incidence (ACLED)			
	Agri Yes	Agri No	Agri50 H	Agri50 L
	(1)	(2)	(3)	(4)
Water Discharge	0.0010	0.0019**	0.0016	0.0010
	(0.0012)	(0.0009)	(0.0015)	(0.0011)
Water Discharge × Shock Down	0.0014**	0.0001	0.0018**	0.0010
	(0.0006)	(0.0019)	(0.0007)	(0.0007)
Water Discharge × Shock Up	-0.0003	0.0000	-0.0005	-0.0008
	(0.0006)	(0.0020)	(0.0008)	(0.0007)
Shock Own	-0.0024	0.0000	-0.0028	-0.0015
	(0.0023)	(0.0013)	(0.0028)	(0.0017)
Shock Down	-0.0009	0.0008	-0.0058	0.0025
	(0.0033)	(0.0016)	(0.0046)	(0.0017)
Shock Up	-0.0010	0.0001	0.0016	-0.0007
	(0.0038)	(0.0016)	(0.0053)	(0.0018)
Cell FE	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓
Dep. Var. Mean	0.11341	0.00995	0.13336	0.03066
R^2	0.41907	0.28129	0.43298	0.33517
Cells	7,124	3,104	5,114	5,114
Observations	178,100	77,600	127,850	127,850

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. In columns (1) and (2) we split the sample according to the presence or absence of agricultural land. In columns (3) and (4) we split the sample according to higher-lower than the median presence of agricultural land. Data for agricultural land are taken from McGuirk and Burke (2020). *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in Section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We expect a higher likelihood of conflict in areas where the returns to accessing water are higher. Given the agrarian nature of the African continent, one of the main ways to exploit water resources is agriculture. To scrutinize this channel we split the sample between cells with high and low level of agricultural ground cover. In particular, in columns 1 and 2 of

Table 3 we split the sample according to whether agriculture is present or totally absent in the cell. On the other hand, in columns 3 and 4 we separate the sample according to the median level of agricultural ground cover. We can detect an impact of downstream shocks only in localities where there is at least a minimum level of agriculture. This finding is in line with these conflicts being over the control of factors for economic production (McGuirk and Burke, 2020). Indeed, the joint presence of water and agricultural land, makes these cells particularly attractive targets for invasion in case of droughts downstream.

Ethnic diversity and cooperation costs

Freshwater resources may be distributed unequally, yet different groups can still cooperate and manage them together. For instance, according to the hydraulic theory, the formation of early states was partly motivated by the necessity of institutions for large-scale irrigation projects (Allen et al., 2020). Moreover, a symbiotic system has often existed between farmers and herders, with herders migrating to farmers' land during dry seasons. This traditional arrangement, especially when farmers' land is situated near rivers, can be seen as a norm that enables efficient sharing of water resources among different groups during periods of limited rainfall. Scarce rainfall in Africa due to climate change threatens established water-sharing institutions, leading to their collapse.

We explore this potential mechanism in Table 4, considering three different measures of imbalance water allocation as detailed in Section 3.2. In columns 1 and 2 we split the sample between cells belonging to neighborhoods with high-low levels of polarization in water access between different ethnic groups. This measure takes the maximum value if in the neighborhood are present two groups owning 50% of the total existing water. The more polarized the access to water is, the higher should be the incentive for groups to appropriate the resource from other populations when they are hit by a shock. As we can observe in Table 4 we only can detect an impact of rainfall shocks in highly polarized neighborhoods. In columns 3 to 6 we do a similar exercise splitting the sample on the basis of two different measures of inequality in water ownership: Gini and Theil indexes. As expected, only shocks happening in markets where inequality in water access is higher have an impact on conflicts incidence. The coefficients corresponding to the interaction *Water Discharge* $\times$ *Shock Down* are way larger and significant in odd columns, indicating that cooperation in water sharing becomes more complex in a context where there is inequality in access to water across different ethnic groups.

Table 4: Ethnic diversity and cooperation costs

	Incidence (ACLED)					
	RQ H	RQ L	Gini H	Gini L	Theil H	Theil L
	(1)	(2)	(3)	(4)	(5)	(6)
Water Discharge	0.0014	0.0016	0.0010	0.0022**	0.0017	0.0019*
	(0.0013)	(0.0015)	(0.0016)	(0.0011)	(0.0016)	(0.0011)
Water Discharge × Shock Down	0.0017***	0.0003	0.0017**	0.0006	0.0018***	0.0008
	(0.0007)	(0.0006)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Water Discharge × Shock Up	-0.0001	-0.0008	-0.0006	-0.0004	-0.0008	-0.0001
	(0.0007)	(0.0006)	(0.0008)	(0.0007)	(0.0008)	(0.0007)
Shock Own	-0.0010	-0.0012	-0.0033	0.0007	-0.0031	0.0007
	(0.0025)	(0.0024)	(0.0027)	(0.0020)	(0.0027)	(0.0019)
Shock Down	-0.0026	0.0028	-0.0029	0.0022	-0.0044	0.0025
	(0.0031)	(0.0023)	(0.0040)	(0.0019)	(0.0039)	(0.0019)
Shock Up	0.0028	-0.0031	0.0022	-0.0018	0.0036	-0.0026
	(0.0034)	(0.0026)	(0.0046)	(0.0022)	(0.0045)	(0.0021)
Cell FE	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓	✓
Dep. Var. Mean	0.08727	0.07869	0.11787	0.04808	0.11907	0.04687
R^2	0.41479	0.44586	0.42949	0.39843	0.42989	0.39681
Cells	5,054	5,052	5,054	5,052	5,054	5,052
Observations	126,350	126,300	126,350	126,300	126,350	126,300

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. In columns (1) and (2) we split the sample according to high-low value of Reynal-Querol polarization index, computed as detailed in Section 3. In columns (3) and (4) we split the sample according to high-low values of Gini index, computed as detailed in Section 3. In columns (5) and (6) we split the sample according to high-low values of the Theil index computed as detailed in Section 3. *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in Section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Water stress

Climate change might generate an increase in conflicts over water resources not just through more frequent droughts, but in the longer run, by depleting the quantity of water present in a given area. Desertification processes are well known to affect some areas of the continent like the Sahel region and more in general, a decrease in water quantity in given areas might break economic equilibria existing among the populations living along river bodies. To explore this possible mechanism, we create a measure of water stress at cell level. In particular, we consider the difference in discharge between the average water presence in a cell during our sample period and the first 10 years for which the variable discharge is available (from 1979 to 1988). Looking at the spatial distribution of the variable (see Figure A8) we can notice how, in most of the continent, there has been a reduction in water quantity over the last 40 years.

In columns 1 and 2 of Table 5 we divide the sample according to higher or lower than the median increase in water presence, while in columns 3 and 4 we split the sample according to a positive or negative change over time in discharge. We can estimate a significant impact of precipitation shocks only in those cells which have experienced a reduction in water presence over time.

Table 5: Water Stress

	Incidence (ACLED)			
	Above Median Change	Below Median Change	Positive Change	Negative Change
	(1)	(2)	(3)	(4)
Water Discharge	0.0017*	-0.0013	0.0019*	0.0007
	(0.0010)	(0.0023)	(0.0010)	(0.0019)
Water Discharge × Shock Down	0.0001	0.0019***	-0.0007	0.0015***
	(0.0007)	(0.0006)	(0.0010)	(0.0005)
Water Discharge × Shock Up	0.0003	-0.0004	0.0012	-0.0005
	(0.0007)	(0.0006)	(0.0010)	(0.0006)
Shock Own	-0.0007	-0.0002	0.0064**	-0.0046**
	(0.0026)	(0.0023)	(0.0028)	(0.0021)
Shock Down	0.0055**	-0.0028	0.0023	-0.0001
	(0.0026)	(0.0026)	(0.0028)	(0.0024)
Shock Up	-0.0004	-0.0023	-0.0025	-0.0007
	(0.0028)	(0.0030)	(0.0028)	(0.0029)
Cell FE	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓
Dep. Var. Mean	0.08827	0.07593	0.07136	0.08813
R^2	0.43971	0.41519	0.42857	0.42647
Cells	5,106	5,105	3,670	6,541
Observations	127,650	127,625	91,750	163,525

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. In columns (1) and (2) we split the sample according to higher-lower than the median water stress as defined in Section 4. In columns (3) and (4) we split the sample according to positive or negative change in water presence between our sample period and the first ten years of water discharge data (1979-1988). *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in Section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Institutions

A key aspect that might ease the consequences of a drought is the ability of the state to re-distribute resources, build infrastructures apt to prevent crises and ensuring property rights protection to avoid violent appropriation of water. In line with the research by Michalopoulos

and Papaioannou (2014), we employ the Worldwide Governance Indicator from the World Bank (Kaufmann et al., 2011) as measures of institutional quality,[10] recognizing its potential significance in facilitating water redistribution under conditions of scarcity. Our analysis primarily considers four key elements: the type of institutional governance, rule of law guarantee, absence of corruption and government effectiveness. In the first two columns of Table 6 we split the sample according to high-low level of democratic governance in a country.[11] We explore whether more democratic systems, characterized by stability and participatory governance, are better equipped to encourage cooperative responses to climate-related challenges. In columns 3 and 4 the focus shifts on high-low levels of rule of law. The idea is that, a better definition and enforcement of property rights are fundamental to managing resources efficiently and resolving disputes, especially in times of environmental stress. In columns 5 and 6 we look into a metric of state capacity, government effectiveness, reflecting the quality of public services and the efficacy of policy formulation and implementation. Higher government effectiveness might contribute to the construction of appropriate infrastructures to cope with climate shocks, but also to respond more rapidly to crises. Lastly, in columns 7 and 8 we split the sample according to high-low levels of corruption. The underlying idea is that property rights protection and government effectiveness necessitate an environment free from corruption. Across all these dimensions, we observe a sizable and significant effect for our primary coefficient of interest only in even columns, indicating countries with weaker institutional quality metrics.

Even if this is mostly correlational evidence and despite we do not have specific data related to effectiveness in water management by states, these results seem to suggest that better institutions might be effective in preventing local violence for water resources in case of climate shocks.

[10] In order to avoid reverse causality issues we consider values of the indexes for the pre-sample period (year 1996).

[11] In particular, we create a measure of democratic governance by computing the mean between "voice and accountability" and "political stability" indexes at country level.

Table 6: Institutional Quality

	Incidence (ACLED)							
	Dem H	Dem L	RLaw H	RLaw L	Gov Eff H	Gov Eff L	Corrupt H	Corrupt L
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Water Discharge	0.0004	0.0016	-0.0007	0.0031	0.0018**	-0.0007	-0.0008	0.0036*
	(0.0008)	(0.0022)	(0.0009)	(0.0019)	(0.0009)	(0.0023)	(0.0009)	(0.0020)
Water Discharge × Shock Down	0.0001	0.0017***	0.0006	0.0016***	0.0005	0.0017***	0.0005	0.0012**
	(0.0006)	(0.0006)	(0.0007)	(0.0006)	(0.0007)	(0.0006)	(0.0006)	(0.0006)
Water Discharge × Shock Up	0.0003	-0.0005	0.0005	-0.0006	-0.0004	0.0002	-0.0003	-0.0001
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0006)	(0.0007)
Shock Own	0.0012	-0.0015	-0.0011	0.0007	0.0036	-0.0033	-0.0015	0.0012
	(0.0023)	(0.0025)	(0.0024)	(0.0024)	(0.0023)	(0.0025)	(0.0020)	(0.0027)
Shock Down	-0.0027	0.0056*	-0.0025	0.0049*	0.0027	-0.0006	-0.0015	0.0052
	(0.0022)	(0.0029)	(0.0024)	(0.0027)	(0.0022)	(0.0030)	(0.0019)	(0.0033)
Shock Up	-0.0041	0.0010	-0.0035	0.0003	0.0015	-0.0050	-0.0017	-0.0008
	(0.0025)	(0.0033)	(0.0027)	(0.0032)	(0.0024)	(0.0034)	(0.0022)	(0.0037)
Cell FE	✓	✓	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Dep. Var. Mean	0.05527	0.11018	0.06696	0.09751	0.06949	0.09575	0.05603	0.10840
R^2	0.36197	0.44512	0.37419	0.45222	0.41544	0.42323	0.41074	0.41853
Cells	5,247	4,981	5,188	5,040	5,351	4,877	5,154	5,074
Observations	131,175	124,525	129,700	126,000	133,775	121,925	128,850	126,850

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. In different columns we split the sample according to higher or lower than the median values in the sample of different variables indicating institutional quality. In particular in columns (1) and (2) we consider democratic governance (which takes into account measures of political stability and voice and accountability), in columns (3) and (4) rule of law, in columns (5) and (6) government effectiveness and in columns (7) and (8) corruption. The indexes are taken from Kaufmann et al. (2011). *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in Section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5. Conclusion

This paper examines the influence of competition for water resources on local violence across the African continent over the period 1997-2021. By combining detailed data on hydrology, river network topology, and weather patterns, we demonstrate that adverse rainfall conditions drive individuals to seek water access in upstream areas with abundant water resources. Our analysis focuses on major conflict events such as battles and violence against civilians. When a downstream cell experiences a rainfall shock, the likelihood of conflict is 0.6 percentage

points larger for a cell with high water presence with respect to a cell where water is scarce. This translates to a 7.30% increase in conflict likelihood with respect to the mean conflict incidence in our sample. Notably, our results remain robust across various "water richness" measures, diverse conflict datasets, and when considering other possible confounders like temperature and population.

Given Africa's predominantly agrarian economy, the economic returns from water access should be higher in areas with significant agricultural output. Consistently with our expectations, we find that the effects are driven by places characterized by higher presence of agriculture.

Additionally, we find that conflict over water resources is more likely in regions characterized by higher cooperation costs, i.e. when water is unevenly distributed across different ethnic groups. Employing three distinct measures of water distribution across ethnicities - polarization, Gini, and Theil indexes - our analysis reveals that greater disparity in water access among different ethnic groups is associated with an increased risk of conflict.

Climate change plays a role not only by increasing the frequency of droughts, but also altering the distribution of surface water, giving rise to desertification processes. We find that the effect is mainly concentrated in those areas where water has been decreasing in the last forty years, possibly destabilizing pre-existing equilibria in terms of water sharing and management.

Finally, we show that institutions, assessed by various World Bank indices, can play a pivotal role in this context. Stronger institutions can mitigate the challenges posed by water scarcity through the development of appropriate infrastructures and the implementation of redistribution schemes. Such strategies can facilitate the equitable allocation of water resources between regions abundant in water and those facing scarcity.

Our results suggest that policymakers should take into account the unequal distribution of freshwater resources when thinking about climate-conflict relationship. Moreover, we highlight the rivers network's structure as an important transmission channel for climate shocks. While we focus on local violence, this structure can shape the relationship between water scarcity and conflict at a larger scale. A prime example is the ongoing geopolitical tensions surrounding the construction of the Grand Ethiopian Renaissance Dam (Climate Diplomacy, 2023*a*). More generally, taking into account the river network is key to understanding how water management policies will affect neighboring regions and countries. This is crucial in a future where climate-related shocks are expected to become increasingly frequent and water scarcity a more acute problem.

Bibliography

Adhvaryu, A., Fenske, J., Khanna, G. and Nyshadham, A. (2021), 'Resources, conflict, and economic development in africa', *Journal of Development Economics* **149**, 102598.

Allen, R. C., Bertazzini, M. C. and Heldring, L. (2020), 'The economic origins of government'.

Almer, C., Laurent-Lucchetti, J. and Oechslin, M. (2017), 'Water scarcity and rioting: Disaggregated evidence from sub-saharan africa', *Journal of Environmental Economics and Management* **86**, 193–209.

Arbatlı, C. E., Ashraf, Q. H., Galor, O. and Klemp, M. (2020), 'Diversity and conflict', *Econometrica* **88**(2), 727–797.

Berman, N., Couttenier, M., Rohner, D. and Thoenig, M. (2017), 'This mine is mine! how minerals fuel conflicts in africa', *American Economic Review* **107**(6), 1564–1610.

Besley, T. and Reynal-Querol, M. (2014), 'The legacy of historical conflict: Evidence from africa', *American Political Science Review* **108**(2), 319–336.

Burke, M., Gong, E. and Jones, K. (2015), 'Income shocks and hiv in africa', *The Economic Journal* **125**(585), 1157–1189.

Burke, M., Hsiang, S. M. and Miguel, E. (2015), 'Climate and conflict'.

Center for International Earth Science Information Network - CIESIN - Columbia University (2018), 'Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11', Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). Accessed on 26-May-2023.
URL: *https://doi.org/10.7927/H4JW8BX5*

Climate Diplomacy (2023*a*), 'Disputes over the grand ethiopian renaissance dam (gerd)', Website.
URL: *https://climate-diplomacy.org/case-studies/disputes-over-grand-ethiopian-renaissance-dam-gerd*

Climate Diplomacy (2023*b*), 'Pastoralist and farmer-herder conflicts in the sahel', Website.
URL: *https://climate-diplomacy.org/case-studies/pastoralist-and-farmer-herder-conflicts-sahel*

Climate Diplomacy (2023*c*), 'Security implications of the gilgel gibe iii dam in ethiopia',

Website.
URL: *https://climate-diplomacy.org/case-studies/security-implications-gilgel-gibe-iii-dam-ethiopia*

Conley, T. G. (1999), 'Gmm estimation with cross sectional dependence', *Journal of econometrics* **92**(1), 1–45.

Corno, L., Hildebrandt, N. and Voena, A. (2020), 'Age of marriage, weather shocks, and the direction of marriage payments', *Econometrica* **88**(3), 879–915.

Depetris-Chauvin, E. (2015), 'State history and contemporary conflict: Evidence from sub-saharan africa', *Available at SSRN 2679594* .

Depetris-Chauvin, E. and Özak, Ö. (2020), 'Borderline disorder:(de facto) historical ethnic borders and contemporary conflict in africa', *Available at SSRN 3541025* .

Döll, P., Kaspar, F. and Lehner, B. (2003), 'A global hydrological model for deriving water availability indicators: model tuning and validation', *Journal of Hydrology* **270**(1-2), 105–134.

Dube, O. and Vargas, J. F. (2013), 'Commodity price shocks and civil conflict: Evidence from colombia', *Review of Economic studies* **80**(4), 1384–1421.

Eberle, U. J. (2020), 'Cep discussion paper no 1694 may 2020 damned by dams? infrastructure and conflict'.

Eberle, U. J., Rohner, D. and Thoenig, M. (2020), 'Heat and hate: Climate security and farmer-herder conflicts in africa'.

Esteban, J., Mayoral, L. and Ray, D. (2012), 'Ethnicity and conflict: An empirical study', *American Economic Review* **102**(4), 1310–1342.

Giuliano, P. and Nunn, N. (2018), 'Ancestral characteristics of modern populations', *Economic History of Developing Regions* **33**(1), 1–17.

Gordon, R. G. and Grimes, B. F. (2009), 'Ethnologue: Languages of the world. 2005', *Dallas, Texas: SIL International, Fifteenth* .

Guariso, A. and Rogall, T. (2017), 'Rainfall inequality, political power, and ethnic conflict in africa'.

Harari, M. and Ferrara, E. L. (2018), 'Conflict, climate, and cells: a disaggregated analysis', *Review of Economics and Statistics* **100**(4), 594–608.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard,

C., Cloke, H. and Pappenberger, F. (2020), 'Glofas-era5 operational global river discharge reanalysis 1979–present', *Earth System Science Data* **12**(3), 2043–2060.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D. and Thépaut, J.-N. (2023), 'ERA5 monthly averaged data on single levels from 1940 to present', Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Accessed on 26-May-2023.

Hsiang, S. M. and Burke, M. (2014), 'Climate, conflict, and social stability: what does the evidence say?', *Climatic change* **123**, 39–55.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2011), 'The worldwide governance indicators: Methodology and analytical issues1', *Hague journal on the rule of law* **3**(2), 220–246.

König, M. D., Rohner, D., Thoenig, M. and Zilibotti, F. (2017), 'Networks in conflict: Theory and evidence from the great war of africa', *Econometrica* **85**(4), 1093–1132.

McGuirk, E. and Burke, M. (2020), 'The economic origins of conflict in africa', *Journal of Political Economy* **128**(10), 3940–3997.

McGuirk, E. F. and Nunn, N. (2020), Transhumant pastoralism, climate change, and conflict in africa, Technical report, National Bureau of Economic Research.

McGuirk, E. F. and Nunn, N. (2021), 'Transhumant pastoralism, climate change and conflict in africa'.

Mekonnen, M. M. and Hoekstra, A. Y. (2016), 'Four billion people facing severe water scarcity', *Science advances* **2**(2), e1500323.

Michalopoulos, S. and Papaioannou, E. (2014), 'National institutions and subnational development in africa', *The Quarterly journal of economics* **129**(1), 151–213.

Michalopoulos, S. and Papaioannou, E. (2016), 'The long-run effects of the scramble for africa', *American Economic Review* **106**(7), 1802–1848.

Miguel, E., Satyanath, S. and Sergenti, E. (2004), 'Economic shocks and civil conflict: An instrumental variables approach', *Journal of political Economy* **112**(4), 725–753.

Monitor, M. E. (2022), 'Egypt farmers fear rising social tensions over scarce water', Website. **URL:** *https://www.middleeastmonitor.com/20221031-egypt-farmers-fear-rising-social-tensions-over-scarce-water/*

Moscona, J., Nunn, N. and Robinson, J. A. (2020), 'Segmentary lineage organization and conflict in sub-saharan africa', *Econometrica* **88**(5), 1999–2036.

Nation (2023), 'Laikipia large-scale farmers accused of diverting rivers into farms', Website.

URL: *https://nation.africa/kenya/counties/laikipia/laikipia-large-scale-farmers-accused-of-diverting-rivers-into-farms-4126256*

Puertas, D. G.-L., van Steenbergen, F., Haile, A. M., Kool, M. and Embaye, T.-a. G. (2021), 'Flood-based farming in africa', PDF.

Raleigh, C., Linke, A., Hegre, H. and Karlsen, J. (2010), 'Introducing acled: an armed conflict location and event dataset: special data feature', *Journal of peace research* **47**(5), 651–660.

Reynal-Querol, M. (2002), 'Ethnicity, political systems, and civil wars', *Journal of conflict resolution* **46**(1), 29–54.

Rohner, D., Thoenig, M. and Zilibotti, F. (2013), 'War signals: A theory of trade, trust, and conflict', *Review of Economic Studies* **80**(3), 1114–1147.

Sarsons, H. (2015), 'Rainfall and conflict: A cautionary tale', *Journal of development Economics* **115**, 62–72.

Strobl, E. and Strobl, R. O. (2011), 'The distributional impact of large dams: Evidence from cropland productivity in africa', *Journal of development Economics* **96**(2), 432–450.

Sundberg, R. and Melander, E. (2013), 'Introducing the ucdp georeferenced event dataset', *Journal of Peace Research* **50**(4), 523–532.

Unfried, K., Kis-Katos, K. and Poser, T. (2022), 'Water scarcity and social conflict', *Journal of Environmental Economics and Management* **113**, 102633.

United Nations (2023), 'UN 2023 Water Conference', https://sdgs.un.org/sites/default/files/2023-05/FINAL%20EDITED%20-%20PGA77%20Summary%20for%20Water%20Conference%202023.pdf.

Verdin, K. L. and Verdin, J. P. (1999), 'A topological system for delineation and codification of the earth's river basins', *Journal of Hydrology* **218**(1-2), 1–12.

World Economic Forum (2023), 'Global risks report 2023', https://www.weforum.org/reports/global-risks-report-2023/digest/.

World Food Programme (2023), 'A women's day on the canal: Green dreams of a farmer amid kenya's drought', Website.
URL: *https://www.wfp.org/stories/womens-day-canal-greens-dreams-farmer-amid-kenyas-drought*

Appendix

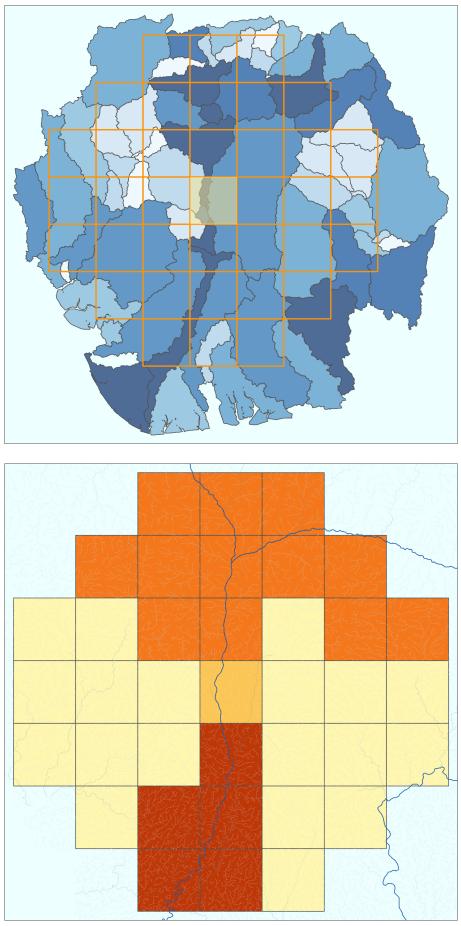
A. Upstream - Downstream

In this section we describe in detail the construction of the rivers network relationships between grid cells sample units. From the hydrology literature (Harrigan et al., 2020), as mentioned in section 3, we take the spatial breakdown of the entire African continent in river basins. A basin can be defined as the area of land drained by a river and its branches. The basins shapefiles are available at different levels of disaggregation; following Eberle (2020) and Strobl and Strobl (2011) we choose level 7 whose basins have an average area comparable to the cells we use in the analysis. Following the Pfafstetter classification system (see Verdin and Verdin, 1999 for a comprehensive explanation about how the system works), for each basin we have information about its position along the river network. In order to understand the relative positioning of our grid cells in terms of up-downstream relationship, we need to assign each cell to a river basin. Given the irregular shape of river basins, there are many different criteria one can use to perform this matching. Since our main objective is to study the interdependence of water resources between different regions, our main criterion to assign a cell to basin is the relative importance in terms of water discharge of the cell's area drained by the basin. In particular, for each intersection between river basins and a given cell we compute the average discharge quantity; then, we assign each cell to the basin whose intersection contains the highest water amount.

We illustrate the methodology by taking as example the confluence of Niger and Benue rivers. In the top panel of Figure A1 we overlay the neighborhood of all the cells whose centroid is within 180 Km from the dark yellow cell at the center of the figure, with the river basins present in the area. The orange grid represents the neighborhood of cells, while the basins are colored on the basis of the average discharge presence. In the bottom panel we display the corresponding assignment of the cells. In light yellow are represented cells whose centroid is located within 180 Km from the reference cell (the cell in the middle in dark yellow) and that do not have any up-downstream relationship with respect to it. The orange (red) cells are those located upstream (downstream) according to our definition. The blue lines represent the rivers with highest water presence in the area (which are indeed the Niger and Benue).

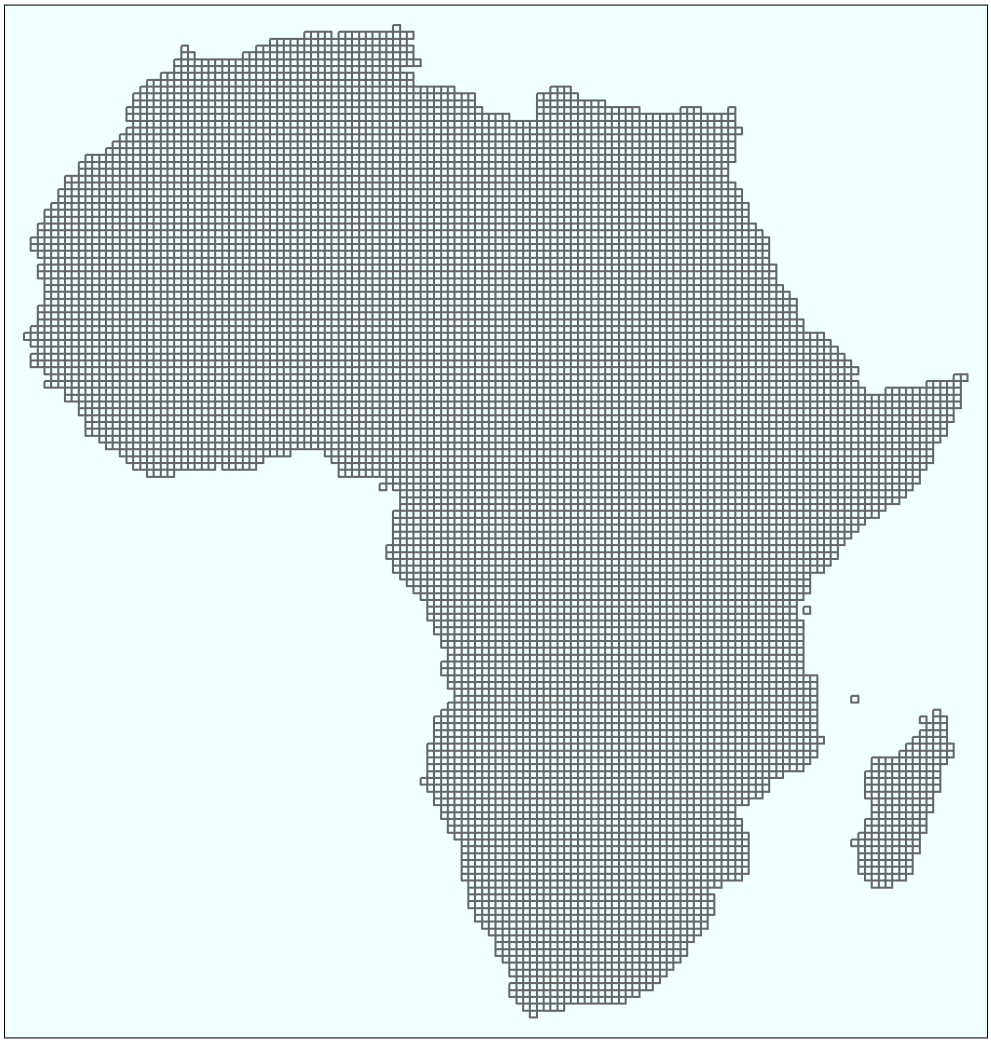
Supplementary Figures

Figure A1: Niger river upstream and downstream relationship



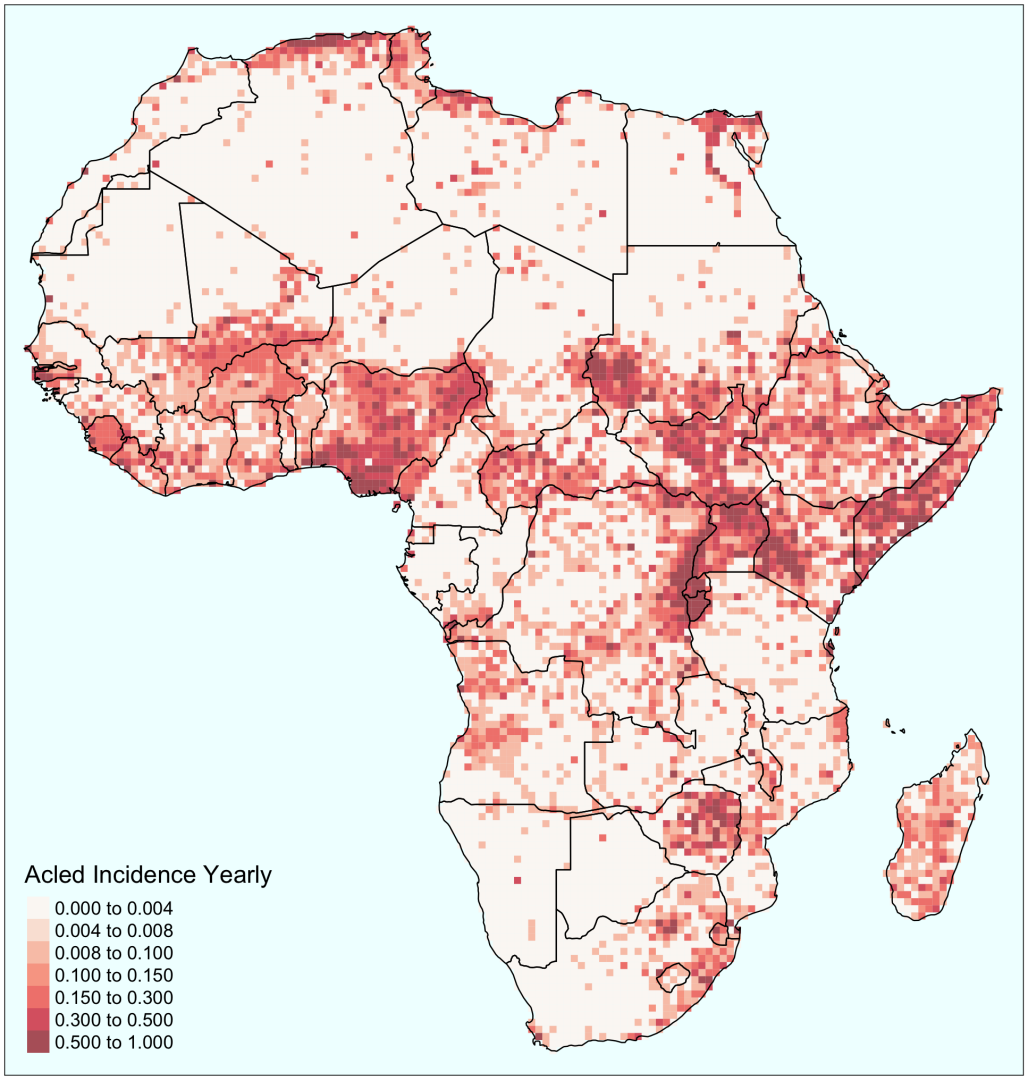
Notes: The figure shows, by way of example, a section of Niger river to illustrate how we build the upstream-downstream relationships. In the top panel we superimpose the grid for the neighborhood (cells within 180 Km radius) of the yellow cell in the center of the figure with the river basins shapefile colored according to the average water discharge present in each of them. In the bottom panel we show the resulting upstream-downstream relationships between the different cells according to the methodology explained in appendix A. Orange cells are those located upstream within the neighborhood of the main cell (in dark yellow), while the red cell are those that we consider downstream with respect to it. Light yellow cells are those coded as neither upstream nor downstream with respect to the main cell.

Figure A2: The grid



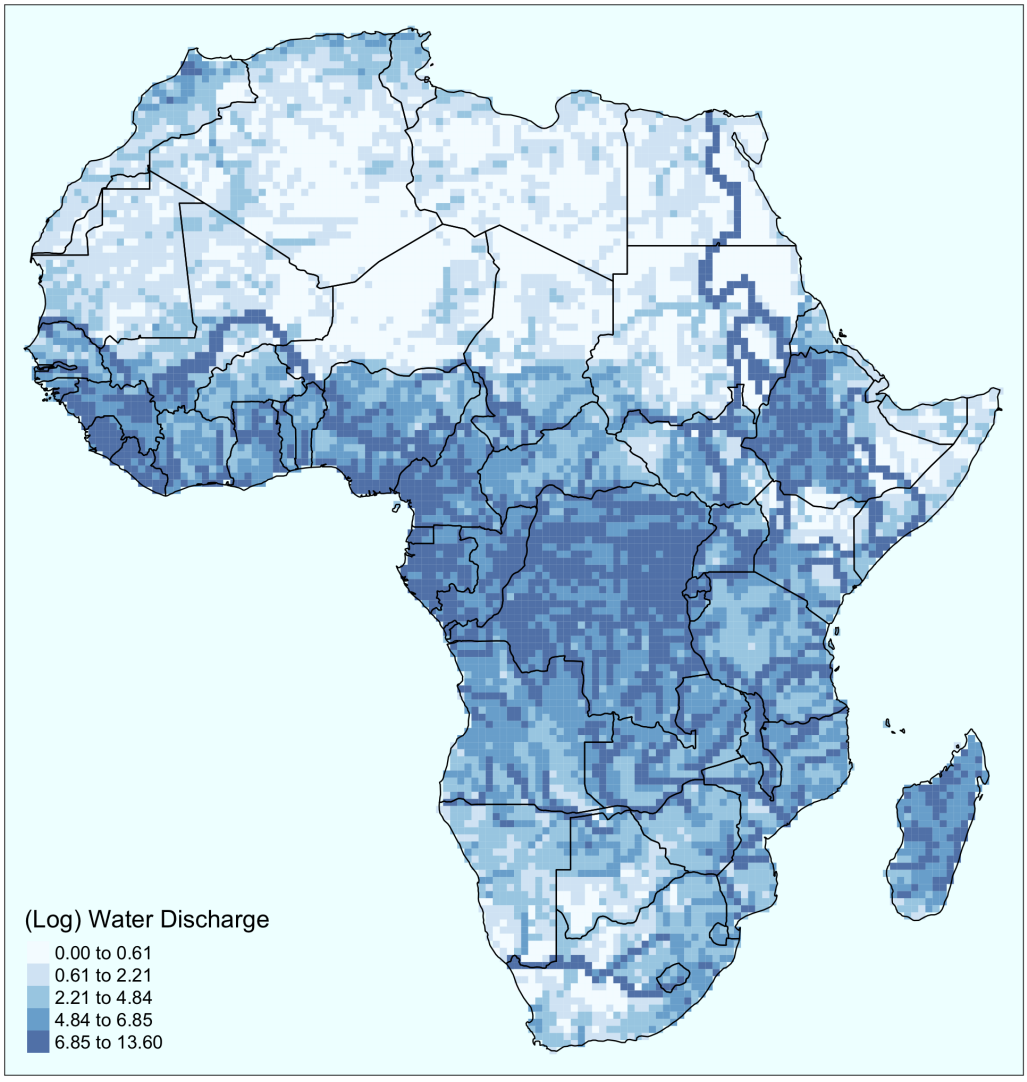
Notes: Grid of $0.5° \times 0.5°$ cells covering the African continent that we use for the analysis.

Figure A3: Conflict (ACLED)



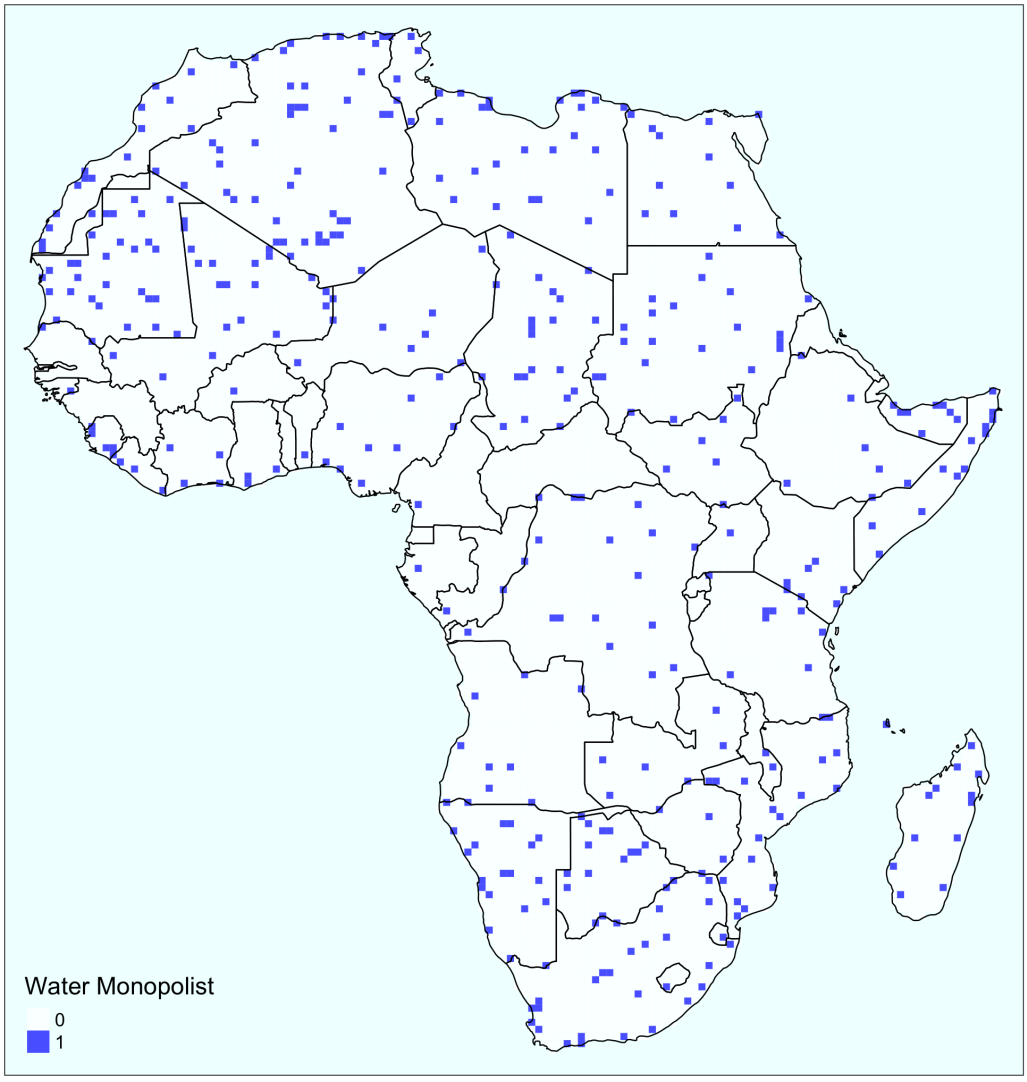
Notes: Spatial distribution of our main dependent variable, conflict incidence, for the period 1997-2021. Darker shadings indicate cells with a higher proportion of years with at least one conflict incident, based on data from the Armed Conflict Location and Event Data Project (ACLED).

Figure A4: Average discharge (cell level)



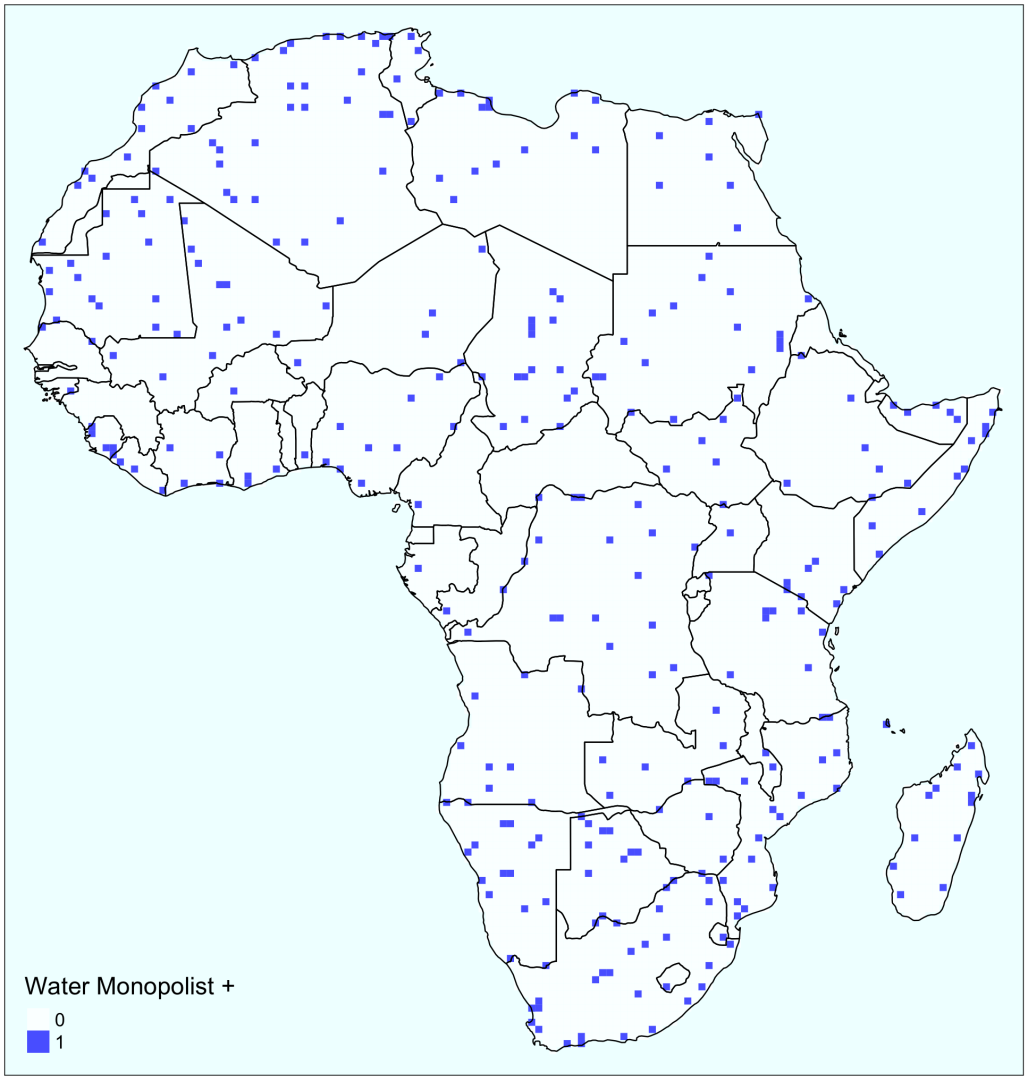
Notes: Cell-level (Log) average yearly discharge in m3/s over the sample period 1997-2021. Darker color indicates areas with higher average discharge. Water discharge data have been taken from Harrigan et al. (2020).

Figure A5: Water Monopolist



Notes: In the map are represented in blue cells which are coded as water monopolist (see Section 3 for details on the definition) for the majority of the years during the sample period 1997-2021.

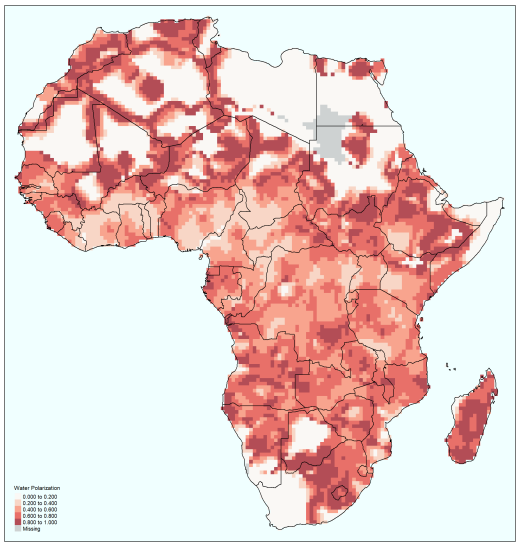
Figure A6: Water Monopolist +



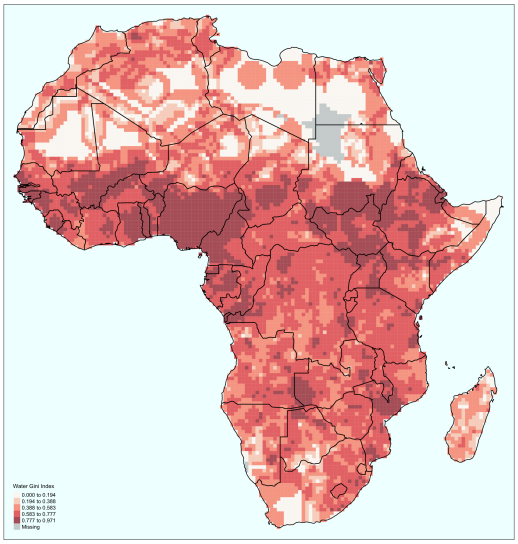
Notes: In the map are represented in blue cells which are coded as water monopolist + (see Section 3 for details on the definition) for the majority of the years during the sample period 1997-2021.

Figure A7: Water Inequality and Polarization

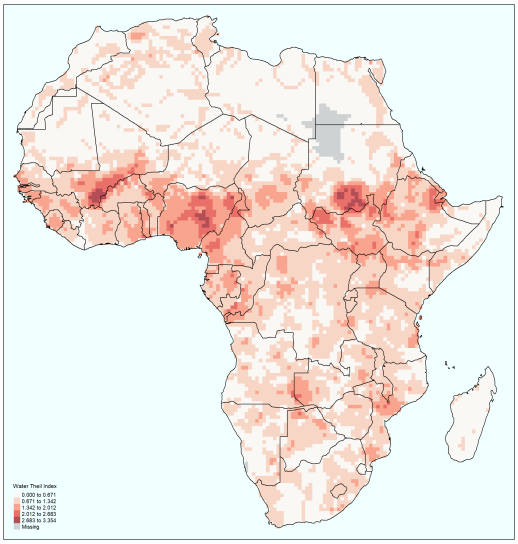
(A) Polarization index



(B) Gini index

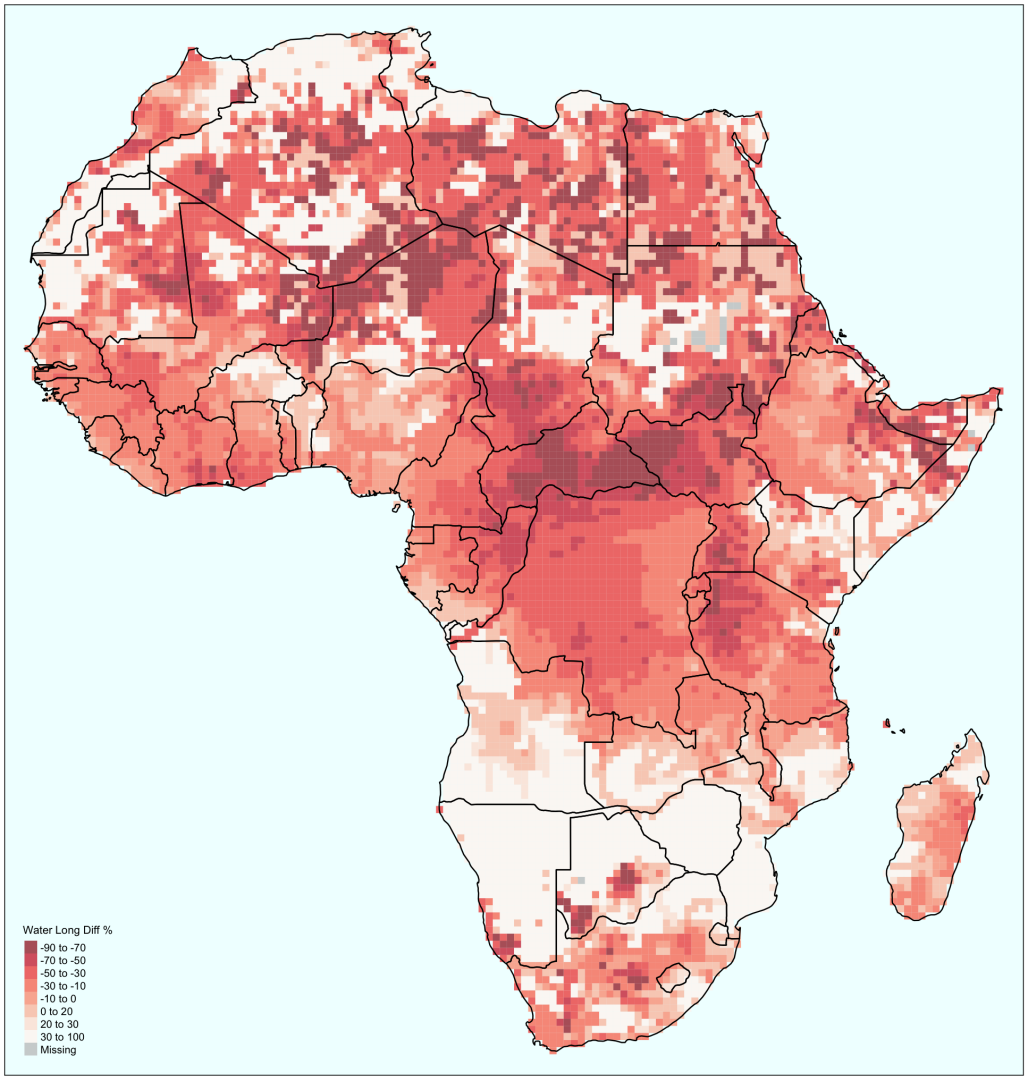


(C) Theil index



Notes: The maps display the spatial distribution of three different measures of water allocation between different ethnic groups at neighborhood level. In top panel (a), we report polarization measure of water ownership, in panel (b) the Gini index, while in panel (c) we report the Theil index distribution. Darker colors indicate higher values of the respective indexes. Grey cells represent missing values.

Figure A8: Water Stress



Notes: The map displays the spatial distribution of the measure of water stress that we use. Darker colors indicate higher level of long term negative changes in water availability. The construction of the measure is detailed in section 4. Discharge data are taken from Harrigan et al. (2020).

Supplementary Tables

Table A1: Summary statistics

Variable	Mean	SD	Min	Median	Max	N
Panel A: *Conflicts*						
Incidence (ACLED)	0.0820	0.2744	0	0	1.0000	255,700
Incidence Battles	0.0543	0.2266	0	0	1.0000	255,700
Incidence Violence	0.0557	0.2293	0	0	1.0000	255,700
Incidence Protests	0.0415	0.1995	0	0	1.0000	255,700
Incidence Riots	0.0351	0.1840	0	0	1.0000	255,700
Incidence (GED)	0.0304	0.1716	0	0	1.0000	337,524
Panel B: *Water measures*						
Water Discharge (ln)	3.6334	3.2151	0	3.0318	14.131	255,700
Water Monopolist	0.0172	0.1299	0	0	1.0000	255,700
Water Monopolist +	0.0125	0.1113	0	0	1.0000	255,700
Panel C: *Rainfall shocks*						
Shock Down	0.2556	0.4362	0	0	1.0000	255,700
Shock Down p10	0.1742	0.3793	0	0	1.0000	255,700
Shock Down p20	0.3337	0.4715	0	0	1.0000	255,700
Shock Own	0.1971	0.3978	0	0	1.0000	255,700
Shock Own p10	0.1273	0.3333	0	0	1.0000	255,700
Shock Own p20	0.2664	0.4421	0	0	1.0000	255,700
Shock Up	0.1755	0.3804	0	0	1.0000	255,700
Shock Up p10	0.1225	0.3278	0	0	1.0000	255,700
Shock Up p20	0.2235	0.4166	0	0	1.0000	255,700
Panel D: *Other variables*						
Agricultural Cover	15.889	24.458	0	2.3642	99.917	255,700
Discharge Long Diff	195.79	2,479.2	-100.00	-16.739	99,670	255,275
Democratic	-0.9225	0.8486	-2.2008	-0.9961	0.9389	255,700
Rule of Law	-0.9089	0.6794	-2.1447	-1.0216	0.5845	255,700
Government Effectiveness	-0.7418	0.6503	-1.9599	-0.9236	1.0205	255,700
Corruption	-0.7347	0.6259	-1.6479	-0.8607	0.8180	255,700
RQ Index	0.5050	0.3053	0	0.5665	1.0000	252,650
Gini Index	0.5614	0.2472	0	0.6171	0.9712	252,650
Theil Index	0.8042	0.5596	0	0.7192	3.3539	252,650
Temperature (day)	27.306	3.4748	10.836	27.239	37.245	255,700
Temperature	24.462	3.4479	8.1089	24.596	34.057	255,700
Population	94,578	317,839	0	20,116	18,604,352	255,700

Notes: The table reports summary statistics for the main variables used in the analysis. The unit of observation is a $0.5^\circ \times 0.5^\circ$ grid cell and year. In *Panel A* we report summary statistics for the measures of conflicts used as dependent variables. In *Panel B* we report the summary statistics for the measures of water presence. In *Panel C* we report summary statistics for the measures of rainfall shocks. Finally, in *Panel D* we report summary statistics for the rest of the variables used for the heterogeneity analysis or as controls.

Table A2: Conley standard errors

	Incidence (ACLED)				
	(1)	(2)	(3)	(4)	(5)
Water Discharge	0.0010	0.0007	0.0010	0.0009	0.0009
	(0.0013)	(0.0013)	(0.0012)	(0.0012)	(0.0012)
Water Discharge × Shock Down	0.0011**		0.0011**		0.0012**
	(0.0006)		(0.0006)		(0.0006)
Water Discharge × Shock Up		0.0003		0.0003	-0.0002
		(0.0006)		(0.0006)	(0.0006)
Shock Down	0.0008		0.0010		0.0009
	(0.0024)		(0.0024)		(0.0023)
Shock Up		-0.0018		-0.0024	-0.0014
		(0.0025)		(0.0025)	(0.0024)
Shock Own			-0.0005	0.0020	0.0000
			(0.0021)	(0.0021)	(0.0021)
Cell FE	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201	0.08201	0.08201
R^2	0.42101	0.42095	0.42101	0.42096	0.42101
Cells	10,228	10,228	10,228	10,228	10,228
Observations	255,700	255,700	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Conley standard errors with a spatial lag of 500 Km and infinite serial correlation are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Alternative data on conflict

	Incidence (GED Geo3)		
	Water Discharge	Water Monopolist	Water Monopolist +
	(1)	(2)	(3)
Water Measure	-0.0003	0.0147**	0.0129*
	(0.0006)	(0.0072)	(0.0070)
Water Measure × Shock Down	0.0005*	0.0143*	0.0305**
	(0.0003)	(0.0086)	(0.0124)
Water Measure × Shock Up	-0.0004	-0.0018	-0.0058
	(0.0003)	(0.0079)	(0.0095)
Shock Own	0.0012	0.0012	0.0012
	(0.0011)	(0.0011)	(0.0011)
Shock Down	0.0001	0.0017*	0.0016*
	(0.0011)	(0.0010)	(0.0010)
Shock Up	-0.0007	-0.0021*	-0.0021**
	(0.0013)	(0.0011)	(0.0011)
Cell FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Dep. Var. Mean	0.03039	0.03039	0.03039
R^2	0.28764	0.28768	0.28771
Cells	10,228	10,228	10,228
Observations	337,524	337,524	337,524

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. Differently from our main analysis we construct the dependent variable using GED dataset. *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1989-2020. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Alternative conflict categories

	Incidence (ACLED) (1)	Incidence Battles (2)	Incidence Violence (3)	Incidence Protests (4)	Incidence Riots (5)
Water Discharge	0.0009	0.0013*	-0.0008	0.0006	-0.0006
	(0.0009)	(0.0008)	(0.0008)	(0.0007)	(0.0006)
Water Discharge × Shock Down	0.0012***	0.0013***	0.0011***	0.0001	0.0001
	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Water Discharge × Shock Up	-0.0002	-0.0003	0.0002	0.0002	0.0005
	(0.0005)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Shock Own	0.0000	0.0004	-0.0011	-0.0011	-0.0030**
	(0.0017)	(0.0015)	(0.0015)	(0.0013)	(0.0012)
Shock Down	0.0009	0.0005	0.0001	-0.0004	-0.0010
	(0.0018)	(0.0016)	(0.0015)	(0.0014)	(0.0012)
Shock Up	-0.0014	-0.0032*	-0.0011	-0.0007	-0.0016
	(0.0021)	(0.0018)	(0.0018)	(0.0016)	(0.0014)
Cell FE	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓
Dep. Var. Mean	0.08201	0.05431	0.05570	0.04152	0.03507
R^2	0.42101	0.36651	0.38268	0.39875	0.37082
Cells	10,228	10,228	10,228	10,228	10,228
Observations	255,700	255,700	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5^\circ \times 0.5^\circ$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one conflict event occurs in a cell and year. In column (1) we report estimates using our main dependent variable which includes ACLED battles and violence against civilians. In columns (2) and (3) we separate the two components of the main dependent variables and consider battles and violence against civilians separately. In columns (4) and (5) we consider less deadly type of conflict events such as protests and riots. *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1989-2020. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Alternative rainfall shocks G10

	Incidence (ACLED)		
	Water Discharge	Water Monopolist	Water Monopolist +
	(1)	(2)	(3)
Water Measure	0.0010	0.0125	0.0163
	(0.0009)	(0.0095)	(0.0104)
Water Measure × Shock Down	0.0014***	0.0275**	0.0458***
	(0.0005)	(0.0137)	(0.0172)
Water Measure × Shock Up	0.0003	-0.0073	-0.0104
	(0.0005)	(0.0120)	(0.0138)
Shock Own	0.0006	0.0002	0.0002
	(0.0020)	(0.0020)	(0.0020)
Shock Down	-0.0003	0.0043**	0.0043**
	(0.0022)	(0.0017)	(0.0017)
Shock Up	-0.0008	0.0009	0.0009
	(0.0024)	(0.0020)	(0.0020)
Cell FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201
R^2	0.42102	0.42101	0.42103
Cells	10,228	10,228	10,228
Observations	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. *Water Measure* indicates generically a measure of water quantity which varies between columns. In column (1) it is the natural logarithm of the average water discharge present in a cell during a given year (*Water Discharge*). In column (2) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year (*Water Monopolist*). In column (3) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year and the discharge is higher than the median level in the sample for that year (*Water Monopolist +*). *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). Differently from the main analysis we define precipitation shocks as precipitation level in a cell-year below the 10th percentile in the long term distribution (see 3 for further details in the construction of precipitation shocks). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Alternative rainfall shocks G20

	Incidence (ACLED)		
	Water Discharge	Water Monopolist	Water Monopolist +
	(1)	(2)	(3)
Water Measure	0.0010	0.0095	0.0115
	(0.0010)	(0.0100)	(0.0108)
Water Measure × Shock Down	0.0013***	0.0218**	0.0386***
	(0.0004)	(0.0104)	(0.0143)
Water Measure × Shock Up	-0.0007	-0.0009	-0.0027
	(0.0005)	(0.0105)	(0.0128)
Shock Own	0.0031**	0.0028*	0.0028*
	(0.0015)	(0.0015)	(0.0015)
Shock Down	-0.0031*	0.0008	0.0008
	(0.0017)	(0.0014)	(0.0014)
Shock Up	0.0003	-0.0023	-0.0023
	(0.0019)	(0.0016)	(0.0016)
Cell FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201
R^2	0.42100	0.42100	0.42103
Cells	10,228	10,228	10,228
Observations	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. *Water Measure* indicates generically a measure of water quantity which varies between columns. In column (1) it is the natural logarithm of the average water discharge present in a cell during a given year (*Water Discharge*). In column (2) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year (*Water Monopolist*). In column (3) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year and the discharge is higher than the median level in the sample for that year (*Water Monopolist +*). *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). Differently from the main analysis we define precipitation shocks as precipitation level in a cell-year below the 20th percentile in the long term distribution (see 3 for further details in the construction of precipitation shocks). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Alternative radius 160 Km

	Incidence (ACLED)		
	Water Discharge	Water Monopolist	Water Monopolist +
	(1)	(2)	(3)
Water Measure	0.0009	0.0102	0.0125
	(0.0009)	(0.0080)	(0.0087)
Water Measure × Shock Down	0.0014***	0.0133	0.0224
	(0.0005)	(0.0109)	(0.0144)
Water Measure × Shock Up	-0.0001	-0.0059	-0.0066
	(0.0005)	(0.0104)	(0.0125)
Shock Own	-0.0001	-0.0005	-0.0005
	(0.0017)	(0.0017)	(0.0017)
Shock Down	0.0002	0.0049***	0.0049***
	(0.0020)	(0.0016)	(0.0016)
Shock Up	-0.0015	-0.0015	-0.0016
	(0.0022)	(0.0018)	(0.0018)
Cell FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201
R^2	0.42102	0.42100	0.42101
Cells	10,228	10,228	10,228
Observations	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. Differently from the main analysis, as robustness exercise, we use 160 Km radius to define a cell neighborhood. *Water Measure* indicates generically a measure of water quantity which varies between columns. In column (1) it is the natural logarithm of the average water discharge present in a cell during a given year (*Water Discharge*). In column (2) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year (*Water Monopolist*). In column (3) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year and the discharge is higher than the median level in the sample for that year (*Water Monopolist +*). *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Alternative radius 200 Km

	Incidence (ACLED)		
	Water Discharge	Water Monopolist	Water Monopolist +
	(1)	(2)	(3)
Water Measure	0.0009	0.0104	0.0093
	(0.0009)	(0.0106)	(0.0112)
Water Measure × Shock Down	0.0010**	0.0217*	0.0341**
	(0.0004)	(0.0123)	(0.0168)
Water Measure × Shock Up	-0.0001	-0.0053	-0.0061
	(0.0005)	(0.0114)	(0.0133)
Shock Own	0.0003	-0.0001	-0.0001
	(0.0017)	(0.0017)	(0.0017)
Shock Down	0.0008	0.0042***	0.0041***
	(0.0018)	(0.0015)	(0.0015)
Shock Up	-0.0018	-0.0018	-0.0019
	(0.0020)	(0.0017)	(0.0017)
Cell FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201
R^2	0.42100	0.42100	0.42101
Cells	10,228	10,228	10,228
Observations	255,700	255,700	255,700

Notes: The table reports estimated coefficients from equation (1). The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. Differently from the main analysis, as robustness exercise, we use 200 Km radius to define a cell neighborhood. *Water Measure* indicates generically a measure of water quantity which varies between columns. In column (1) it is the natural logarithm of the average water discharge present in a cell during a given year (*Water Discharge*). In column (2) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year (*Water Monopolist*). In column (3) it is an indicator variable equal to 1 if the cell is the one with the highest water discharge in a neighborhood in a given year and the discharge is higher than the median level in the sample for that year (*Water Monopolist +*). *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Additional Controls

	Incidence (ACLED)			
	(1)	(2)	(3)	(4)
Water Discharge	0.0009	0.0013	0.0015	0.0008
	(0.0009)	(0.0010)	(0.0010)	(0.0009)
Water Discharge × Shock Down	0.0012***	0.0012***	0.0012***	0.0011**
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Water Discharge × Shock Up	-0.0001	-0.0002	-0.0002	-0.0002
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Shock Own	0.0000	-0.0004	-0.0006	-0.0002
	(0.0017)	(0.0017)	(0.0017)	(0.0017)
Shock Down	0.0009	0.0007	0.0005	0.0004
	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Shock Up	-0.0014	-0.0015	-0.0015	-0.0002
	(0.0021)	(0.0021)	(0.0021)	(0.0020)
Log pop.	0.0046			
	(0.0048)			
Temp.		0.0044**		
		(0.0022)		
Temp. (day)			0.0059***	
			(0.0020)	
Lagged Incidence				0.1701***
				(0.0051)
Cell FE	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓
Dep. Var. Mean	0.08201	0.08201	0.08201	0.08366
R^2	0.42102	0.42103	0.42104	0.44153
Cells	10,228	10,228	10,228	10,228
Observations	255,700	255,700	255,700	245,472

Notes: The table reports estimated coefficients from equation (1) with additional controls. The unit of observation is a $0.5° \times 0.5°$ grid cell and year. The dependent variable is a dummy that takes value 1 if at least one violent conflict occurs in a cell and year. *Water Discharge* is the natural logarithm of the average water discharge present in a cell during a given year. *Shock* is an indicator variable taking value 1 if a location experiences a drought (as defined in section 3), upstream (*Up*), downstream (*Down*) or within the unit of observation (*Own*). In different columns we introduce additional controls to our baseline regression equation. In particular in column (1) we control fro (Log) population in the cell, in column (2) we controls for average temperature over the year, in column (3) we control for average daily temperature over the year and in column (4) we control for conflicts happening in the previous year. The sample covers the years in the interval 1997-2021. Clustered standard errors by cell are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

CHAPTER 4

Price and Prejudice: Housing Rents Reveal Racial Animus

Marius Brülhart — University of Lausanne
Gian-Paolo Klinke — Federal Statistical Office
Andrea Marcucci — University of Lausanne
Dominic Rohner — University of Lausanne
Mathias Thoenig — University of Lausanne

Abstract

We study market housing rents in neighborhoods of asylum seeker hosting centers that vary by their ethnic composition. We exploit the opening and closure of centers and the quasi-random spatial allocation of asylum seekers in Switzerland. Rents within 0.7km of an active center are found on average to be 3.8% lower than in the control group. The price drop is significantly more pronounced when centers host a higher share of asylum seekers from Sub-Saharan countries. However, neither religious affiliation nor inferred crime propensity affect prices significantly. Our findings are consistent with racial animus as main driver of observed market outcomes.

Keywords: ethnic prejudice, willingness to pay, housing prices, refugee centers.

JEL: D90, J15, R31.

Introduction

Xenophobia and discrimination are global social challenges. While voluntary international migration is economically efficient, it may give rise to political backlash, especially among host societies. For example in the United Kingdom, survey respondents systematically named immigration as the country's most important issue in the run-up to the Brexit vote in 2016 (Blinder and Richards, 2016). In the United States, immigration and/or race relations regularly feature very prominently among problems mentioned in nationwide polls, taking top spot during periods of heightened immigration pressure in 2014, 2018 and 2024 (Jones, 2024).

The reasons for anti-migrant sentiment can be manifold. One factor may be the perception that migrants are competing against natives in labor and housing markets. The resulting distributional conflict can be exacerbated by non-economic animus against people of different nationality or ethnicity. Such prejudice – the object of our paper – has been amply documented through laboratory experiments, field experiments and observational studies.[1]

Yet, not everybody is prejudiced. In a population of heterogeneous types, it is uncertain *a priori* whether and how much prejudice will matter at the aggregate level. Models of labor market discrimination, for example, show that non-prejudiced employers will arbitrage away the biases of prejudiced employers, so that prejudice may not affect the aggregate market outcome (Becker, 1957; Heckman, 1998). In the presence of market frictions, however, arbitrage will be incomplete, and discriminatory preferences will to some extent be reflected in market prices (Black, 1995). Moreover, prejudice has been shown to be a rather weakly held preference. People who express prejudiced opinions in unincentivized surveys or in choices among otherwise equivalent alternatives may not act on their prejudice when discrimination incurs a cost. In an incentivized field experiment, Hedegaard and Tyran (2018) have found that the probability of ethnically discriminating falls by 9 percent for every 10 percent rise in the price of doing so. For both those reasons – coexistence of unequally prejudiced agents, and cost sensitivity of discriminatory behavior – real-world market outcomes could conceivably reveal no discrimination even if a nonzero share of market participants hold preferences that are consistent with prejudice in weakly incentivized settings.

Our aim is to measure such an aggregate-level equilibrium outcome in a market with the likely presence of prejudiced agents. For this, we need to be able to observe actual market prices in a setting that features measurable and plausibly exogenous changes in the scope for ethnic discrimination. Our approach is to track the evolution of housing prices in the neighborhood of state-run asylum seeker hosting centers (henceforth "asylum centers") with

[1] See Neumark (2018) and Lang and Kahn-Lang Spitzer (2020) for recent surveys.

different occupant populations. The opening of a center typically represents a salient and quantitatively relevant increase in the local-level population of various foreign origins. Asylum seekers, however, do not have access to local labor and housing markets, which means that they do not compete directly with residents. Hence, any resulting changes in local housing prices plausibly reflect the market equilibrium outcome determined by prejudiced and non-prejudiced natives. Moreover, center openings are driven by determinants outside the affected local community and are thus largely exogenous with respect to local conditions. Housing price movements in the vicinity of asylum centers therefore offer us a measure of the equilibrium market response to immigration. A unique feature of our research design is that data on compositional differences in the populations of asylum centers also allow us to explore market reactions to different immigrant types.

Specifically, our analysis draws on geo-coded data for (a) the universe of public residential rental postings in Switzerland and (b) the opening, closing and populations of asylum centers over the 2004-2014 period. We estimate the effect of non-vacant centers on local rental prices applying a comprehensive set of fixed effects to filter out time-invariant confounders. The identifying variation we consider stems from changes around the opening and closing of asylum centers, comparing housing units within the same neighborhood but located at different distances from the center, and by comparing centers with differing occupant populations. We find that the opening of a center is on average associated with a drop of 3.8% in rental prices in close proximity to the center (within 0.7km). This effect emerges immediately after the opening of a center and persists for at least two years after the opening. To investigate the underlying mechanisms, we exploit a unique feature of our setting: the quasi-random allocation of asylum seekers across centers, implying exogenous differences in the nationality composition of different centers (Couttenier, Petrencu, Rohner and Thoenig, 2019). Unlike related previous studies, we observe those differences in the data. We find that the drop in local rental prices in the vicinity of centers populated by above-median shares of asylum seekers from Sub-Saharan Africa equals 4.9%, whereas around centers with below-median shares of Sub-Saharan Africans it equals 2.7%. Sub-Saharan African origin is the only asylum-seeker characteristic we find to have a statistically significant impact on housing-price responses. This result persists when controlling for crime-related variables, suggesting that the drops in rental prices may at least in part be due to animus other than "statistical" discrimination based on crime propensities.

This paper is related to several strands of existing research. First, we contribute to an empirical literature studying discrimination of ethnic minorities in various contexts, such as the labor market (Lang and Lehmann, 2012; Agan and Starr, 2018; Neumark, 2018; Hangartner, Kopp and Siegenthaler, 2021), the housing market (Yinger, 1986; Ewens, Tomlin and Wang, 2014; Laouénan and Rathelot, 2022), citizenship applications (Hainmueller and Hangartner, 2013),

or criminal justice (Lang and Kahn-Lang Spitzer, 2020). With regard to asylum seekers, there is evidence from a large non-incentivized survey that asylum seekers with greater employability, more consistent asylum testimonies, severe vulnerabilities and of Christian (rather than Muslim) faith are met with greater public acceptance in European host countries (Bansak, Hainmueller and Hangartner, 2016).

Methodologically, we follow the hedonic pricing approach, using granular housing-market data to infer willingness to pay. This method has in the past been used to gauge the impact of a range of factors. In Table A1, we list the most closely related studies we are aware of, focusing on the price effects of factors that one may expect to be considered as disamenities.[2] The upper part of the table summarizes papers that have explored the effects of asylum centers. Our average estimated effect is close to the midpoint of those found for other countries. Switzerland therefore does not appear to exhibit particularly pronounced housing market responses to asylum centers. What sets our paper apart is that we draw on newly assembled, very fine-grained data on the asylum center nationality composition and on local and group-specific crime rates, which enables us to relate price effects to center population characteristics. In the lower part of the table, we list papers that use closely comparable estimation methods but consider other types of local disamenity. Again, our central elasticity estimate is within the range of those found elsewhere for disamenities including noise, toxic sites, and the presence of sex, gun or drugs offenders.

Table A1 also shows that, in contrast to our approach, most prior studies use housing prices (rather than rents) with the idea of quantifying the capitalized effect of *permanent* disamenities. Instead, following Saiz (2003) and Saiz (2007), we use housing rents. In our context, using rents is more appropriate given the transitory nature of the amenity shocks we are interested in and the fact that rents respond to market conditions more quickly than housing prices (Rosen and Smith, 1983). Indeed, we look not only at short-lived openings of asylum centers but also at their ethnic composition, which varies over time – often from one month to another.[3]

The rest of the paper is organized as follows: Section 1 describes the method and data, Section 2 displays the main results, and Section 3 concludes. The Appendix contains further context and data description and additional robustness checks.

---

[2] A key distinction to keep in mind is that while asylum centers may yield merely subjective disamenities for certain (prejudiced) individuals, several other forms of exposure listed in Table A1 lead to objective negative externalities affecting all individuals.

[3] Further related literatures study the impact of asylum seeker arrival on crime (Bianchi, Buonanno and Pinotti, 2012; Bell, Fasani and Machin, 2013; Couttenier et al., 2019), on right-wing voting (Otto and Steinhardt, 2014; Barone, D'Ignazio, De Blasio and Naticchioni, 2016; Steinmayr, 2016; Dustmann, Vasiljeva and Piil Damm, 2019) and on policy preferences (Zimmermann and Stutzer, 2022). For another literature linking the population composition to urban outcomes, see Eberle, Henderson, Rohner and Schmidheiny (2020).

1. Method and Data

1.1 Estimation

We estimate the effect of asylum centers on local rental prices using a difference-in-differences strategy, where housing units located near an asylum center (treatment group) are compared to units located further away (control group), before and after the opening or, less frequently, closing of the center. Our basic observational unit, h, is a rental posting. In order to contrast housing units that are comparable in terms of local economic and topographic characteristics, we only consider dwellings within a two-kilometer radius of asylum centers. We assign each unit h to the closest asylum center, c, that is open at some point during the time range covered by our sample. Moreover, we control for a vector of housing characteristics, $\mathbf{H}_h$, a vector of center fixed effects, λ_c, a vector of municipality fixed effects, γ_m, and a vector of year fixed effects, $\tau_{y[t]}$.

Formally we estimate the following equation:

$$\begin{aligned}\ln(Rent)_{hcmt} = \alpha + \beta(Active_{hct} \times Prox_{hc}) + \theta_c Active_{hct} + \delta_c Prox_{hc} \\ + \mathbf{H}_h^{'}\mathbf{\Gamma} + \lambda_c + \gamma_m + \tau_{y[t]} + \varepsilon_{hcmt},\end{aligned} \tag{1}$$

where the dependent variable is the natural logarithm of the rental price per square meter in rental posting h, assigned to its nearest asylum center c, located in municipality m, and published on day t. The binary variable $Active_{hct}$ is set to one if the center c assigned to rental posting h is open at time t when the property advert is recorded. The binary variable $Prox_{hc}$ is set to one if the property referred to by rental posting h is located within a certain threshold distance from its nearest center c. Distance thresholds are algorithm-driven, as detailed below.

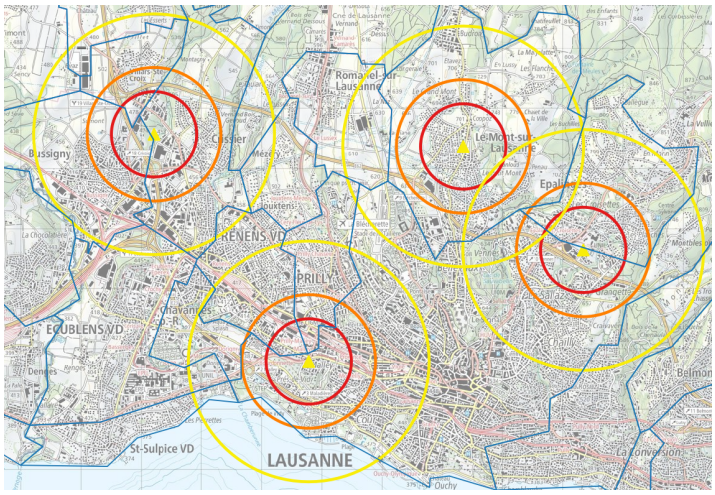
Notice that because of the fixed-effect structure in equation (1) and the inclusion of the center-specific coefficients θ_c and δ_c, we exploit only within-center variation to estimate β, our parameter of interest.[4] This approach is important for two reasons. First, it allows us to contrast housing units located in comparable neighborhoods. Second, it reduces the concern that the "forbidden-comparisons" typical of staggered settings between already treated and newly treated units could bias our coefficients (e.g. Borusyak, Jaravel and Spiess, 2021). To further ease this concern we present a replication of our main estimates using an alternative stacked regression specification technique in Appendix C.

Our treatment and control groups are illustrated in Figure 1. Yellow rings show the 2-kilometer circles around asylum centers which delineate our estimation sample. Treated hous-

[4]This is due to the inclusion of three sets of fixed effects: center-specific fixed effects, λ_c, center-specific dummy variables for housing units observed during the open spell of a center, $\theta_c Active_{ct}$, and center-specific dummy variables for housing units within the treatment radius, $\delta_c Prox_{hc}$.

ing units are those located within the red rings (i.e. below the retained threshold distance). Properties located between the red and the orange circles are excluded from the sample in order to avoid capturing spillover effects.[5] Moreover, we drop from the sample housing units located at the same time within the treated (red) ring of a center and the treatment (red) or spillover (orange) ring of their second closest center, in order to avoid contamination of the estimated effects by centers' opening and closing at different points in time.[6] We determine the length of the treatment (red) and spillover (orange) radii using a method proposed by Butts (2023) to estimate treatment effect decay as a function of distance non-parametrically, exploiting the partitioning-based least squares approach developed by Cattaneo, Farrell and Feng (2020). In Panel 2A of Figure 2, we show that the strongest effect is present until 709 meters from the centers. Then it attenuates from 710 to 1,110 meters, to become even smaller between 1,110 and 1,425 meters. We choose as the baseline spillover ring the 710–1,110 meter band, in order to avoid losing too many observations, but as can be noticed in column (2) of Panel 2B, results are robust if we drop also the 1,110–1,425 meter band.

Figure 1: Asylum Centers and their Neighborhoods



Notes: The figure shows, by way of example, the four asylum centers in the urban area of Lausanne with the radii for the respective treatment and control groups. Blue lines are municipality borders. The underlying map is obtained from https://www.swisstopo.admin.ch/it/geodata/maps/smr/smr50.html.

[5] We can think of two kinds of spillover effects. One is the negative effect on housing rents that might propagate further away than our treatment radius, as Figure 2A suggests. The second type of spillover effect is individuals moving away from the treatment area after a center's opening, re-settling in the control group region and thus increasing the rental prices there.

[6] Imagine a house located in the intersection area of the red circle and the orange circle, respectively, of two different centers A and B. Center A is closer to the housing unit than center B (by definition). If we observe a rental posting for the unit when center A is still closed but center B is already open, then we would get a biased estimate considering the house as untreated while it might be affected by the fact that center B is already open.

1.2 Data

1.2.1 Asylum Centers and Surrounding Housing

Upon their arrival in Switzerland, asylum seekers are required to register at one of seven federal registration centers located at the main border crossings and airports, where identity checks and first interviews take place. There, asylum seekers are assigned to one of the twenty-six cantons according to an exogenous allocation scheme based on population size (see discussion in Couttenier et al., 2019). The cantonal authorities are then responsible for the distribution of asylum seekers among municipalities, and they decide on the opening of new centers and where to locate them.

We obtained non-publicly available data on asylum-seeker hosting centers from 13 cantonal authorities or, in some cases, from private bodies mandated by the cantons to manage the centers. The data include the date of opening and, where relevant, closing, the precise location, and the hosting capacity for all centers which were opened at some point in time over the 2004-2014 period. As we seek to retain those asylum centers that are plausibly salient to the local population, we consider only centers with a capacity of at least 30 beds and that stay open for at least four years within our sample period.

We draw on a database of internet advertisements for rental housing units in Switzerland over the period 2004-2014. The dataset originates from on-line publications on 26 national and regional web platforms and has been provided by meta-sys.ch, a consulting firm. Our sample consists of a repeated cross section of 157,708 housing units with an average of about 13 transactions per center neighborhood and month over our sample period. 82% of our units of observation correspond to single floor apartments, and the average annualized rent per square meter was 280 Swiss francs (CHF, with 1 CHF $\approx$ 1 USD). Further summary statistics about housing units in the sample can be found in Table A2. Among the information available in the dataset, we rely on the publication date of the offer, the geo-coordinates of the housing unit, its rental price and its floor space. We also include dummy variables for the number of rooms and a set of 11 categories that define the type of housing unit. Throughout the analysis we take housing price to be the annualized rent, in CHF, per square meter. We match housing units to asylum centers on distance "as the crow flies" regardless of the activity status of the center.

1.2.2 Asylum Seekers

We have access to non-publicly available individual-level administrative data for all asylum seekers arriving in Switzerland during our sample period (see Couttenier et al., 2019). We match this information to hosting centers, based on the address and time. In this way we are able to retrieve the nationality of origin and the stated religion of individuals living in each

hosting center at any given time. These data also allow us to construct shares of different types of asylum center populations, based on nationality and religion.

As an additional variable to characterize asylum center populations, we consider the average genetic distance between nationalities, as constructed by Spolaore and Wacziarg (2018). Given that phenotype and genotype are imperfectly correlated, this offers an alternative measure of "otherness". We retain the genetic distance of asylum seekers living in the centers with respect to the *overall* population of Swiss nationals and compute the time-varying weighted average of genetic distance for every asylum center ($Gendist$). Furthermore, we compute a time-invariant measure of *local* genetic distance between the (time averaged) nationalities of host municipality and asylum center populations ($Local\ Gendist$).[7]

1.2.3 Crime Data

We also have access to non-publicly available data on all crimes detected by the police in Switzerland between 2009 and 2014 (see Couttenier et al., 2019). The data include precise information about perpetrators' nationality and residence status. We exploit this information to distinguish crimes committed by asylum seekers and build a measure of "crime propensity" by nationality. The measure is constructed by dividing the total number of crimes committed by asylum seekers of a given nationality over the total number of asylum seekers from that nationality living in Switzerland. We then build a center-level measure of crime propensity, by taking a weighted average of national crime propensities of individuals from different countries present in a center on a given day. We validate this measure in Appendix Table A4, which confirms that it is effectively correlated with the municipality-month-level number of crimes committed by asylum seekers.

2. Results

Figure 2 presents our baseline estimates of equation (1). Panel 2A illustrates non-parametric estimates of treatment effects in five bins with an equal number of observations.[8] It is apparent that the effect of an open asylum center on housing rents is negative in the immediate vicinity and dissipates with distance. We take this evidence to guide our baseline choice of radii to define our treatment and control groups, defining as the treatment group all housing units within a 709 meter radius of a given center and as the control group all housing units

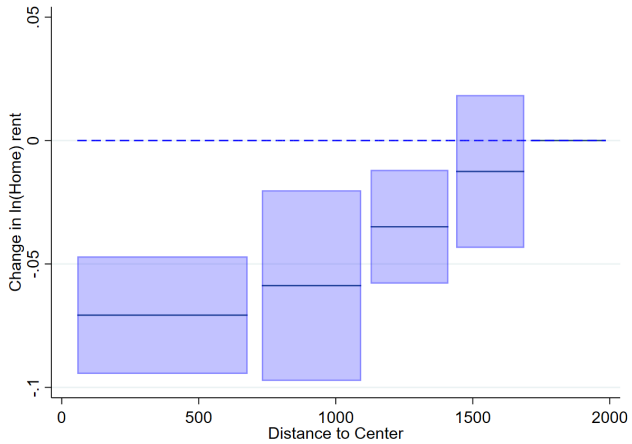
[7] More precisely, we construct the weighted sum of the genetic distances among all nationalities present in the municipality and all nationalities present in the center. The weights are given by the product of time-averaged nationality shares in the local population and among center residents.

[8] Specifically, we follow the approach developed by Butts (2023), splitting the sample into distance quantiles following Cattaneo et al. (2020). The effect is then estimated non parametrically within each bin, comparing units pre and post treatment. The estimated effect from the most distant bin of observations is then subtracted from the others by way of normalization.

within the corresponding 1,110–2,000 meter distance band.

Figure 2: Average Effect on Rental Prices of Asylum Centers

(A) Distance Gradient



(B) Baseline Estimates

	(1)	(2)	(3)	(4)
Dep. variable	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)
Radii in meters:	*baseline*			
treatment; spillover	709; 1,110	709; 1,425	500; 1,000	500; 1,500
$Active \times Prox$	-0.0383***	-0.0299**	-0.0508***	-0.0428**
	(0.0124)	(0.0115)	(0.0174)	(0.0183)
Observations	154,708	106,833	152,730	78,277
R^2	0.453	0.446	0.456	0.449

Notes: In Panel (a), we plot the non-parametric estimates of the treatment effect as a function of distance, following Butts (2023). In Panel (b), we report the estimated coefficient β from equation (1). The unit of observation is a housing unit advert h published on day t. The sample covers observations within 2 kilometers of an asylum center that was open for at least four years within the period 2004-2014 and had a hosting capacity of at least 30 beds. In each column we control for a set of housing characteristics (described in Section 1.2.1), year and municipality fixed effects, as well as the fixed effects to exploit within center variation $\theta_c Active_{hct}$, $\delta_c Prox_{hc}$ and λ_c in accordance with equation (1). Clustered standard errors by municipality are reported in parentheses. The total number of municipalities (clusters) present in the sample is 192. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

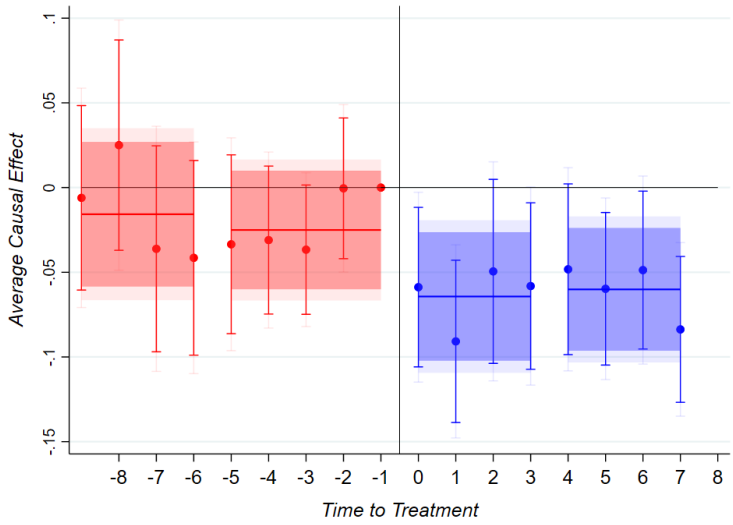
In Panel 2B we show difference-in-differences estimates of the price effects for varying radii. Column (1) applies our baseline definition of treated, spillover and control groups. According to our baseline estimate, the opening of an asylum center reduces average rental prices in the vicinity by 3.8 percent.[9] Columns (2)-(4) show that this estimate is robust to changing

[9]We systematically report coefficient estimates $\hat{\beta}$ rather than $(e^{\hat{\beta}} - 1)$, as the difference is negligible within the value ranges that we obtain.

the definition of the treatment and spillover groups. In column (2), we increase the spillover radius to 1,425 meters. Doing so we exclude from the sample also all those observations which are in the third bin (from left to right) in Panel 2A. Finally, in columns (3)-(4) we show that the result does not significantly change when we apply round distance cutoffs at 500, 1,000 and 1,500 meters. Decreasing the treatment radius from 709 to 500 meters results in a somewhat larger coefficient of -5.1 percent (column 3), which is however not statistically significantly different from our baseline estimate.

The difference-in-differences estimates presented in Figure 2 examine rental prices over the pre- versus post-opening periods. Taking an event-study approach, we can study the time profile of price changes and test for pre-trends. In particular, given the structure of our dataset with some centers closing and re-opening over time, we use the methodology developed in De Chaisemartin and D'Haultfoeuille (2022).[10] We consider quarterly time intervals two years prior to and two years after the opening of asylum centers, focusing on the baseline treatment and control definitions. Those estimates are shown in Figure 3.

Figure 3: Time Pattern of Rental Prices Around the Opening of Asylum Centers



Notes: The figure displays the event-study estimates of the main effect $Active_{hct} \times Prox_{hc}$, at quarterly frequency for two years prior and two years after the opening of an asylum center. The estimates are computed using the approach proposed by De Chaisemartin and D'Haultfoeuille (2022).

While the price series are somewhat volatile, the negative effect of center openings on rental

[10] As far as we are aware, among the estimation methods recently developed (see Borusyak et al., 2021, Callaway and Sant'Anna, 2021, and Sun and Abraham, 2021) to estimate leads and lags of treatment effects, correcting the potential bias due to negative weights, this is the only one allowing for a setting where the treatment can switch on and off over time.

prices in the immediate neighborhood again emerges clearly. Interestingly, that effect appears already in the first quarter subsequent to the opening of a center. This suggests that the rental price effect of asylum centers does not build up gradually, as the practical impacts of a center's presence get noted in the neighborhood, but happens immediately upon the activation of a center. We do not find statistically significant evidence of any pre-trends in local rental prices, which supports the interpretation of our estimates as causal effects of asylum centers on rental prices.

Next, we add interaction effects to our difference-in-differences regressions in order to explore whether different asylum center populations generate different rental price effects. We consider two types of heterogeneity. One approach is to consider the "crime propensity" of asylum seekers, as described in Section 1.2.3. We take this as a variable that could proxy for statistical discrimination, whereby rental price movements might reflect observed or latent changes in local crime risks due to the presence of asylum seekers. Our alternative approach is to consider simple socio-ethnic distinctions: religion, average genetic distance from Swiss natives, and skin color. We take such variables as potential proxies for prejudice (sometimes also referred to as animus or taste-based discrimination).

We present the main results in Table 1, with complementary estimates shown in the Appendix (Tables A7, A8; Figure A1). When we interact the baseline effect with a dummy variable set to one for centers whose populations at a given time have above-median inferred crime propensity, we find only a borderline significant negative effect. The total effect obtained by summing the double and triple interactions of $Crime$ is not statistically different from zero. We detect a considerably larger and more precisely estimated difference when we split asylum centers into those with above-median and below-median shares of residents of Sub-Saharan African nationality. According to the estimates of column (3) in Table 1, the opening of a "low-African" center reduces local rental prices by 2.7 percent, but that of a "high-African" center reduces them by 4.8 percent. When we consider both interactions jointly (column 4 of Table 1), the interaction with $Crime$ is insignificant while that with $African$ remains statistically significant and quantitatively large.

As we show in Appendix Table A7, religion, defined as the share of center residents who self-declare as Muslim, does not appear to drive differences in rental-price responses. The variable $Gendist$ that captures genetic distance relative to the overall population of Swiss nationals, however, has a similarly strong effect as that of the share of Sub-Saharan Africans – which is not surprising, given that the two variables have a correlation coefficient of 0.90. Overall, our estimates suggest that the skin color of center residents is the main source of heterogeneous rental price responses. This is consistent with racial prejudice rather than

statistical discrimination.[11]

Table 1: Socio-Ethnic Differences in Center Populations

	(1)	(2)	(3)	(4)
Dep. Variable	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)
Effect	Base	Crime	African	African & Crime
$Active \times Prox$	-0.0383***	-0.0314**	-0.0266**	-0.0239**
	(0.0124)	(0.0129)	(0.0107)	(0.0108)
$Active \times Crime$		0.0067		0.0065
		(0.0045)		(0.0051)
$Active \times Prox \times Crime$		-0.0107*		-0.0058
		(0.0057)		(0.0059)
$Active \times African$			0.0008	-0.0006
			(0.0084)	(0.0083)
$Active \times Prox \times African$			-0.0221***	-0.0203***
			(0.0049)	(0.0059)
Observations	154,708	154,708	154,708	154,708
R^2	0.453	0.453	0.453	0.453

Notes: The table reports estimated coefficients β from equation (1) as well as interaction effects with socio-demographic variables representing the inferred crime propensity of a center's population ($Crime$) and the of Sub-Saharan African presence in a given center ($African$). $Crime$ and $African$ are binary variables set to one whenever the population of a given center has an above-median crime propensity (inferred from nationalities, see Appendix B) and an above-median share of Sub-Saharan African people, respectively, with relevant medians calculated over the entire data sample. The unit of observation is a rental posting, $h = hcmt$. The sample covers observations within 2 kilometers of an asylum center that was open for at least four years within the period 2004-2014 and had a hosting capacity of at least 30 beds. In each column we control for a set of housing characteristics (described in Section 1.2.1), year and municipality fixed effects, as well as the fixed effects to exploit within center variation $\theta_c Active_{hct}$, $\delta_c Prox_{hc}$ and λ_c in accordance with equation (1)). Clustered standard errors by municipality are reported in parentheses. The number of sample municipalities (clusters) is 192. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, we investigate whether rental market price effects differ across localities with different characteristics. We split the sample by the median across municipalities of one of two variables: the average education level of the local municipality's population, and the local

[11] Note that in our data Sub-Saharan origin and crime propensity do not correlate. In fact, the crime propensity of asylum seekers from Sub-Saharan Africa is one-third lower than the crime propensity of non-Sub-Saharan asylum seekers. This difference, however, is not statistically significant (p-value $= 0.34$). Appendix Table A8 shows that the effect of $African$ is robust to varying the cutoff values for generating the binary variable. Appendix Figure A1, based on a nonparametric estimation explained in Appendix D, suggests the effect of $African$ to be nonlinear, with notable discontinuity above the 60th percentile, which corresponds to an African share of 28%. This implies a threshold effect and likely explains why replacing $African$ by its continuous version does not yield statistically significant estimates on the triple interaction term.

genetic distance between municipality and center populations. As these characteristics are available only at the level of municipalities, we assign to the housing units within 2km from a center the characteristics pertaining to the municipality where the center itself is located. The education variable can serve to proxy for the "education hypothesis", whereby ethnic and national prejudice diminishes with exposure to formal education (see, e.g., Dustmann and Preston, 2007; Hainmueller and Hiscox, 2007). In the same vein, if rental prices respond less sensitively to the opening of an asylum center in municipalities where the local population is less dissimilar from asylum seekers hosted in the center, this could be (loosely) interpreted as consistent with the "contact hypothesis" (see, e.g., Allport, Clark and Pettigrew, 1954; Pettigrew and Tropp, 2006; Rohner, Thoenig and Zilibotti, 2013; Rohner and Zhuravskaya, 2023), in the sense that areas that are already ethnically diverse *before* the opening of an asylum center, are better used to interacting with immigrants of diverse national origins.[12] Our identification strategy here exploits the cross-municipality variations in the local genetic distance variable (see Table A2). This approach relies on the fact that Switzerland is a small country that hosts a diverse and large population of immigrants. Hence, Swiss municipalities differ remarkably in terms of the nationalities of origin of their residents.

Results are shown in Table 2. Consistent with the education hypothesis, we find that rental prices in municipalities with above-median educational attainment react somewhat less strongly to asylum centers than those in municipalities with below-median educational attainment (columns 1 and 2), although the difference is not statistically significant. When we split the sample according to local genetic distance (columns 3 and 4), rental prices seem to respond slightly more sensitively to the opening of an asylum center in municipalities where the local population is more dissimilar from asylum seekers hosted in the center, in line with the (loosely defined) contact hypothesis. This difference, however, is not statistically significant either.

[12] Note that this corresponds to a broad interpretation of the contact hypothesis, since in its narrow definition this hypothesis only applies to contact between specific groups. Put differently, in the standard formulation of the contact hypothesis, more contact between people from, say, Switzerland and Senegal would not affect Swiss attitudes towards people from a third country, say, Mali. Also in the game-theoretic micro foundations of Rohner et al., 2013, the trust building effect of peaceful interaction is confined to matching between two specific groups, and does not give rise of generalized open-mindedness towards other groups.

Table 2: Differences in the Composition of Local Resident Populations

	(1)	(2)	(3)	(4)
Dep. Variable	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)
	Education		*Local Gendist*	
	High	Low	High	Low
$Active \times Prox$	-0.0300**	-0.0519***	-0.0442**	-0.0332
	(0.0128)	(0.0193)	(0.0179)	(0.0201)
Observations	79,581	74,866	79,765	73,813
Clusters	57	142	107	100
R^2	0.330	0.429	0.426	0.479
T-test (p)	.345	.345	0.709	0.709

Notes: The table reports estimated coefficients β_1 from equation (1). The unit of observation is a rental posting, $h = hcmt$. The sample covers rental postings for housing units located within 2km of a hosting center, for the period 2004-2014. We split the total sample by two municipal characteristics: $Education$, defined as the share of residents with either a university degree or a higher professional qualification ("école professionnelle supérieure"), and local genetic distance ($Local\ Gendist$). In even-numbered (odd-numbered) columns, the sample is composed of observations in municipalities with below-median (above-median) values of the given municipal characteristic. In each column we control for a set of housing characteristics (described in Section 1.2.1), year and municipality fixed effects, as well as the fixed effects to exploit within center variation $\theta_c Active_{hct}$, $\delta_c Prox_{hc}$ and λ_c in accordance with equation (1). In the bottom row of the table, we report p-values of two sided t-tests for the equality of the double interaction $Active \times Prox$. Clustered standard errors by municipality are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3. Summary and Discussion

Does individual-level ethnic prejudice affect aggregate market outcomes, or are such biases arbitraged away? To answer this question, we investigate equilibrium prices in a setting featuring the likely presence of prejudiced agents: real estate transactions in the neighborhood of asylum seeker hosting centers. We employ difference-in-differences estimation of rental housing prices, exploiting the fact that asylum seekers in Switzerland are allocated across centers quasi-randomly. Market rents of housing units within 0.7km of an active center are found on average to be 3.8% lower than market rents in the control group, for at least two years (the length of our observation window in the event-study analysis). Arbitrage by non-prejudiced agents is therefore partial at best.

The price drop varies markedly with the share of asylum seekers of Sub-Saharan African origin: in the vicinity of centers with below-median Sub-Saharan African shares, the price effect is -2.7%, but for centers with above-median shares, that effect is -4.8%. In contrast, we find no statistically significant effect heterogeneity with respect to inferred crime propensity

or religious affiliation of asylum seekers. Those findings are consistent with racial animus as the dominant driver of the observed market outcomes. We also find suggestive evidence consistent with the education and contact hypotheses: the estimated rental price effect of center opening is somewhat stronger in municipalities where local population is less educated and more diverse in terms of genetic distance from asylum seekers hosted in the center.

How large are these effects? Our estimate of the impact of an asylum center is quantitatively comparable to what is typically found in the hedonic pricing literature regarding the impact of perceived criminality at the local level (e.g. presence of sex offenders, narcotics labs).[13]

Another way of benchmarking our estimated effect is to calculate the implied equivalent (monetary) variation. In our sample, the average yearly rent per square meter is 280 CHF, and the average surface is equal to 80 sqm (see Appendix Table A2). Hence, the average drop of 3.8% in the local rental price implies that the yearly willingness to pay in order to avoid an asylum center opening nearby is estimated to be 851 CHF per year and housing unit ($=0.038 \times 280 \times 80$), or some 0.7% of average gross household income (SFSO, 2011). The implied willingness to pay for an asylum center's Sub-Saharan African population share to drop from above-median to below-median is 492 CHF or some 0.4% of income.

We can cross-validate our main estimate against an external data source by considering the case of the Swiss municipality of Oberwil-Lieli (population $\approx 2,300$), where in 2016 a majority of citizens voted in favour of collectively paying a fine of 110 CHF per day and asylum seeker for *not* hosting asylum seekers assigned to their municipality by the allocation rule. A back-of-the envelope calculation shows that the willingness to pay as elicited by this vote is actually comparable to the one based on our empirical analysis of housing-market price responses.[14]

An important general lesson from the current paper is that there is a "price of prejudice". Similar to the well-known result of Dal Bó and Dal Bó (2011) that criminal and appropriative activities make everybody worse off, including the criminal, racial prejudice imposes costs not only on the objects of prejudice but also on the subjects, as shown in our computations of willingness to pay. This means that policies promoting tolerance and open-mindedness are win-win and feature a double dividend – for both victims and discriminators. A grow-

[13] For a summary of comparable estimates, see Table A1.

[14] The minimum hosting capacity of asylum centers considered in our study is 30 beds. Hence, for preventing the opening of such an asylum center, the citizens of Oberwil-Lieli would be ready to pay 1,204,500 CHF per year ($= 110 \times 30 \times 365$). For a population of some 2,300, this translates into 524 CHF per citizen per year. The representative housing unit in Switzerland hosts 2.2 individuals (SFSO, 2023); hence we obtain a willingness to pay of 1,152 CHF per year and housing unit. If instead we take the the mean center capacity of 95 (see Table A Panel A), this amount even increases to 3,648 CHF per year and housing unit. Such an extrapolation is probably unrealistic, given that the actual number of asylum seekers assigned to that municipality was 10. Moreover, Oberwil-Lieli is a rural and politically conservative place. Nonetheless, the outcome of that uniquely informative local referendum suggests that our estimated willingness to pay of 851 CHF per year and housing unit is not implausibly high.

ing literature studies how inter-group tolerance can be fostered, drawing on the theoretical premises of the contact hypothesis, according to which more frequent (fair and peaceful) interactions with people of different ethnic and national background reduce prejudice (see, e.g., Allport et al., 1954; Pettigrew and Tropp, 2006; Rohner et al., 2013; Rohner and Zhuravskaya, 2023). Recent empirical evidence from national (military) service (Samii, 2013; Okunogbe, 2018; Cáceres-Delpiano, De Moragas, Facchini and González, 2021), soccer (Mousa, 2020; Alrababa'h, Marble, Mousa and Siegel, 2021) or reconciliation ceremonies (Cilliers, Dube and Siddiqi, 2016) show that inter-group interaction can foster inter-group trust and reduce tensions. A promising avenue for future research would be to explore the effect of inter-group interaction on racial animus towards asylum seekers more closely.

Bibliography

Agan, A. and Starr, S. (2018), 'Ban the box, criminal records, and racial discrimination: A field experiment', *Quarterly Journal of Economics* **133**(1), 191–235.

Allport, G. W., Clark, K. and Pettigrew, T. (1954), *The Nature of Prejudice*, Addison-Wesley, Reading, MA.

Alrababa'h, A., Marble, W., Mousa, S. and Siegel, A. (2021), 'Can exposure to celebrities reduce prejudice? The effect of Mohamed Salah on islamophobic behaviors and attitudes', *American Political Science Review* **115**(4), 1111–1128.

Bansak, K., Hainmueller, J. and Hangartner, D. (2016), 'How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers', *Science* **354**(6309), 217–222.

Barone, G., D'Ignazio, A., De Blasio, G. and Naticchioni, P. (2016), 'Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives' voting behavior', *Journal of Public Economics* **136**, 1–13.

Becker, G. S. (1957), *The Economics of Discrimination*, University of Chicago Press.

Bell, B., Fasani, F. and Machin, S. (2013), 'Crime and immigration: Evidence from large immigrant waves', *Review of Economics and statistics* **95**(4), 1278–1290.

Bianchi, M., Buonanno, P. and Pinotti, P. (2012), 'Do immigrants cause crime?', *Journal of the European Economic Association* **10**(6), 1318–1347.

Black, D. A. (1995), 'Discrimination in an equilibrium search model', *Journal of Labor Economics* **13**(2), 309–334.

Blinder, S. and Richards, L. (2016), 'UK public opinion toward immigration: Overall attitudes and level of concern', *Migration Observatory briefing, COMPAS, University of Oxford* .

Boes, S. and Nüesch, S. (2011), 'Quasi-experimental evidence on the effect of aircraft noise on apartment rents', *Journal of Urban Economics* **69**(2), 196–204.

Borusyak, K., Jaravel, X. and Spiess, J. (2021), 'Revisiting event study designs: Robust and efficient estimation', *arXiv preprint arXiv:2108.12419* .

Butts, K. (2023), 'Difference-in-differences with geocoded microdata', *Journal of Urban Eco-*

nomics: Insights **133**(103493).

Cáceres-Delpiano, J., De Moragas, A.-I., Facchini, G. and González, I. (2021), 'Intergroup contact and nation building: Evidence from military service in Spain', *Journal of Public Economics* **201**, 104477.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

Cattaneo, M. D., Farrell, M. H. and Feng, Y. (2020), 'Large sample properties of partitioning-based series estimators', *Annals of Statistics* **48**(3), 1718–1741.

Cengiz, D., Dube, A., Lindner, A. and Zipperer, B. (2019), 'The effect of minimum wages on low-wage jobs', *Quarterly Journal of Economics* **134**(3), 1405–1454.

Cilliers, J., Dube, O. and Siddiqi, B. (2016), 'Reconciling after civil conflict increases social capital but decreases individual well-being', *Science* **352**(6287), 787–794.

Congdon-Hohman, J. M. (2013), 'The lasting effects of crime: The relationship of discovered methamphetamine laboratories and home values', *Regional Science and Urban Economics* **43**(1), 31–41.

Couttenier, M., Petrencu, V., Rohner, D. and Thoenig, M. (2019), 'The violent legacy of conflict: Evidence on asylum seekers, crime, and public policy in Switzerland', *American Economic Review* **109**(12), 4378–4425.

Currie, J., Davis, L., Greenstone, M. and Walker, R. (2015), 'Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings', *American Economic Review* **105**(2), 678–709.

Daams, M. N., Proietti, P. and Veneri, P. (2019), 'The effect of asylum seeker reception centers on nearby house prices: Evidence from the Netherlands', *Journal of Housing Economics* **46**, 101658.

Dal Bó, E. and Dal Bó, P. (2011), 'Workers, warriors, and criminals: social conflict in general equilibrium', *Journal of the European Economic Association* **9**(4), 646–677.

De Chaisemartin, C. and D'Haultfoeuille, X. (2022), Difference-in-differences estimators of intertemporal treatment effects, Technical report, National Bureau of Economic Research.

Dealy, B. C., Horn, B. P. and Berrens, R. P. (2017), 'The impact of clandestine methamphetamine labs on property values: Discovery, decontamination and stigma', *Journal of Urban Economics* **99**, 161–172.

Deshpande, M. and Li, Y. (2019), 'Who is screened out? Application costs and the targeting

of disability programs', *American Economic Journal: Economic Policy* **11**(4), 213–48.

Dröes, M. I. and Koster, H. R. (2023), 'A world divided: Refugee centers, house prices and household preferences', *Journal of Economic Geography* **23**(1), 51–90.

Dustmann, C. and Preston, I. P. (2007), 'Racial and economic factors in attitudes to immigration', *BE Journal of Economic Analysis & Policy* **7**(1).

Dustmann, C., Vasiljeva, K. and Piil Damm, A. (2019), 'Refugee migration and electoral outcomes', *Review of Economic Studies* **86**(5), 2035–2091.

Eberle, U. J., Henderson, J. V., Rohner, D. and Schmidheiny, K. (2020), 'Ethnolinguistic diversity and urban agglomeration', *Proceedings of the National Academy of Sciences* **117**(28), 16250–16257.

Ewens, M., Tomlin, B. and Wang, L. C. (2014), 'Statistical discrimination or prejudice? a large sample field experiment', *Review of Economics and Statistics* **96**(1), 119–134.

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277.

Guignet, D., Jenkins, R. R., Nolte, C. and Belke, J. (2023), 'The external costs of industrial chemical accidents: A nationwide property value study', *Journal of Housing Economics* **62**, 101954.

Hainmueller, J. and Hangartner, D. (2013), 'Who gets a Swiss passport? a natural experiment in immigrant discrimination', *American Political Science Review* **107**(1), 159–187.

Hainmueller, J. and Hiscox, M. J. (2007), 'Educated preferences: Explaining attitudes toward immigration in europe', *International Organization* **61**(2), 399–442.

Hangartner, D., Kopp, D. and Siegenthaler, M. (2021), 'Monitoring hiring discrimination through online recruitment platforms', *Nature* **589**(7843), 572–576.

Heckman, J. J. (1998), 'Detecting discrimination', *Journal of Economic Perspectives* **12**(2), 101–116.

Hedegaard, M. S. and Tyran, J.-R. (2018), 'The price of prejudice', *American Economic Journal: Applied Economics* **10**(1), 40–63.

Hennig, J. (2021), 'Neighborhood quality and opposition to immigration: Evidence from German refugee shelters', *Journal of Development Economics* **150**, 102604.

Jones, J. M. (2024), 'Immigration surges to top of Most Important Problem list', *Gallup (https://news.gallup.com/poll/611135/immigration-surges-top-important-problem-list.aspx, accessed 15/03/2024)* .

Lang, K. and Kahn-Lang Spitzer, A. (2020), 'Race discrimination: An economic perspective', *Journal of Economic Perspectives* **34**(2), 68–89.

Lang, K. and Lehmann, J.-Y. K. (2012), 'Racial discrimination in the labor market: Theory and empirics', *Journal of Economic Literature* **50**(4), 959–1006.

Laouénan, M. and Rathelot, R. (2022), 'Can information reduce ethnic discrimination? Evidence from Airbnb', *American Economic Journal: Applied Economics* **14**(1), 107–32.

Liang, J. and Alexeev, S. (2022), 'Harm reduction or amplification? the adverse impact of a supervised injection room on housing prices', *Regional Science and Urban Economics* **98**(103856).

Linden, L. and Rockoff, J. E. (2008), 'Estimates of the impact of crime risk on property values from Megan's Laws', *American Economic Review* **98**(3), 1103–1127.

Mousa, S. (2020), 'Building social cohesion between christians and muslims through soccer in post-isis iraq', *Science* **369**(6505), 866–870.

Muehlenbachs, L., Spiller, E. and Timmins, C. (2015), 'The housing market impacts of shale gas development', *American Economic Review* **105**(12), 3633–59.

Muñoz-Morales, J. and Singh, R. (2023), 'Do school shootings erode property values?', *Regional Science and Urban Economics* **98**, 103852.

Neumark, D. (2018), 'Experimental research on labor market discrimination', *Journal of Economic Literature* **56**(3), 799–866.

Okunogbe, O. M. (2018), 'Does exposure to other ethnic regions promote national integration? evidence from Nigeria', *World Bank Policy Research Working Paper* (8606).

Otto, A. H. and Steinhardt, M. F. (2014), 'Immigration and election outcomes: Evidence from city districts in Hamburg', *Regional Science and Urban Economics* **45**, 67–79.

Pettigrew, T. F. and Tropp, L. R. (2006), 'A meta-analytic test of intergroup contact theory.', *Journal of Personality and Social Psychology* **90**(5), 751.

Pope, J. C. (2008*a*), 'Buyer information and the hedonic: The impact of a seller disclosure on the implicit price for airport noise', *Journal of Urban Economics* **63**(2), 498–516.

Pope, J. C. (2008*b*), 'Fear of crime and housing prices: Household reactions to sex offender registries', *Journal of Urban Economics* **64**(3), 601–614.

Rohner, D., Thoenig, M. and Zilibotti, F. (2013), 'War signals: A theory of trade, trust, and conflict', *Review of Economic Studies* **80**(3), 1114–1147.

Rohner, D. and Zhuravskaya, E. (2023), 'Nation building: Big lessons from successes and failures'.

Rosen, K. T. and Smith, L. B. (1983), 'The price-adjustment process for rental housing and the natural vacancy rate', *American Economic Review* **73**(4), 779–786.

Saiz, A. (2003), 'Room in the kitchen for the melting pot: Immigration and rental prices', *Review of Economics and Statistics* **85**(3), 502–521.

Saiz, A. (2007), 'Immigration and housing rents in American cities', *Journal of Urban Economics* **61**(2), 345–371.

Samii, C. (2013), 'Perils or promise of ethnic integration? Evidence from a hard case in Burundi', *American Political Science Review* **107**(3), 558–573.

SFSO (2011), 'Haushaltbudgeterhebung 2009 (Survey of Household Finances)', Swiss Federal Statistical Office.

SFSO (2023), 'Number of occupants per dwelling', Swiss Federal Statistical Office. *www.bfs.admin.ch/bfs/en/home/statistics/construction-housing/dwellings/housing-conditions/persons.html (accessed 02/02/2023)*.

Spolaore, E. and Wacziarg, R. (2018), 'Ancestry and development: New evidence', *Journal of Applied Econometrics* **33**(5), 748–762.

Steinmayr, A. (2016), 'Exposure to refugees and voting for the far-right: (unexpected) results from Austria', IZA Discussion Paper, 9790.

Sun, L. and Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.

Tang, C. K. and Le, T. (2023), 'Crime risk and housing values: Evidence from the gun offender registry', *Journal of Urban Economics* **134**, 103526.

Taylor, L. O., Phaneuf, D. J. and Liu, X. (2016), 'Disentangling property value impacts of environmental contamination from locally undesirable land uses: Implications for measuring post-cleanup stigma', *Journal of Urban Economics* **93**, 85–98.

Van Vuuren, A., Kjellander, J. and Nilsson, V. (2019), 'Refugees and apartment prices: A case study to investigate the attitudes of home buyers', *Regional Science and Urban Economics* **77**, 20–37.

Yinger, J. (1986), 'Measuring racial discrimination with fair housing audits: Caught in the act', *American Economic Review* pp. 881–893.

Zimmermann, S. and Stutzer, A. (2022), 'The consequences of hosting asylum seekers for

citizens' policy preferences', *European Journal of Political Economy* **73**, 102130.

Appendix

This Appendix complements the paper "Price and Prejudice: Housing Rents Reveal Racial Animus".

A. Summary Statistics

In Table A2 we report summary statistics for our sample of rental postings and hosting centers. Panel A lists characteristics of housing units advertised in our sample rental postings, Panel B lists variables relating to the status of the nearest hosting center by rental posting, and in Panel C we summarize the data on capacity and open spells of our sample hosting centers.

B. Crime Propensity Measure: Validation

In this Appendix we validate the measure of crime propensity used as our proxy for statistical discrimination. One possible issue is that, differently from $African$, our taste-based discrimination measure that is clearly visible for the local population, residents nearby an asylum center may not really realize if asylum seekers living there are more or less crime-prone.
To verify that our crime propensity variable proxies for the number of crimes committed locally by asylum seekers, we estimate the following Poisson regression model:

$$\ln \mathbb{E}(\lambda_{ms}) = \beta_1 Active_{ms} + \beta_2(Active_{ms} \times \ln Capacity_{ms}) + \beta_3(Active_{ms} \times Crime_{ms}) + \mathbf{X}'_{ms}\mathbf{\Gamma} + \tau_m + \gamma_{y[s]} + \delta_{k[s]}. \quad (2)$$

We report the estimates of the model in Table A4. The units of observation are municipalities, m, in which at least one asylum center of our sample is present. The dependent variable, $\ln \lambda_{ms}$, is the log number of crimes committed by asylum seekers (odd columns) or by the total population (even columns) in each municipality and month (s). The time span is determined by the availability of crime data (2009-2014). The variable $Active_{ms}$ takes the value of 1 if at least one center is active in municipality m and month s. The variable $Capacity_{ms}$ corresponds to the total installed capacity in the centers opened in municipality m during month s. The variable $Crime_{ms}$ takes the value of 1 if the asylum seekers' crime propensity in the center in municipality m during month s is higher than the median value in the sample. In the vector $\mathbf{X}_{ms}$ we gather other controls such as the natural logarithm of population in municipality m at time s and the interaction between center activity and the high African presence dummy, $Active \times African$, which we control for in columns (5) to (8) of Table A4. τ_m is a municipality fixed effect, $\gamma_{y[s]}$ is a year fixed effect, and , $\delta_{k[s]}$ is a calendar-month fixed effect (to filter out potential seasonality in criminal activity).

Our main coefficients of interest is β_3, which captures the effect of having crime-prone asylum seekers hosted in the center(s) located in a given municipality. The coefficient is positive and statistically significant only when we have as dependent variable the number of crimes perpetrated by asylum seekers. This confirms that our measure of crime propensity is a good proxy for statistical discrimination. Yet, one can notice that high presence of Sub-Saharan African individuals does not have any statistically significant effect on overall local crime.

C. Stacked Regression

As an alternative to our main specification (equation 1), in order to ease potential concerns that negative weights may affect and bias our main results, we replicate Figure 2 (Panel B) and Table 1 estimates using a stacked difference-in-differences approach (see Cengiz, Dube, Lindner and Zipperer 2019, and Deshpande and Li 2019). Even if with the fixed effect structure in equation 1 we compare housing units within each center, for those centers that open and close multiple times in our sample period (9 out of 91) we have the treatment event of opening happening at different points in time. This might bias our estimates in the case treatment effects are heterogeneous over time.[15] More precisely, the model we use is:

$$\begin{aligned}\ln(Rent)_{hemt} = \tilde{\alpha} + \tilde{\beta}(Active_{het} \times Prox_{he}) + \mathbf{H}_h'\mathbf{\Gamma} + \tilde{\theta}_e Active_{het} \\ + \tilde{\delta}_e Prox_{he} + \tilde{\lambda_e} + \tilde{\gamma}_m + \tilde{\tau}_{y[t]} + \tilde{\varepsilon}_{hemt}.\end{aligned} \tag{3}$$

The difference between equations 1 and 3 is that in the latter we define an event e as the opening of a center and we include an event-specific fixed effects structure. In this way, for those centers opening and closing over our sample period we are implicitly defining different treatment and control groups of rental postings, stacking opening events and avoiding "forbidden comparisons" between already treated and newly treated housing units. For those centers which open (close) one time during our sample period, the units of observations' assignment to the event (opening or closing) coincides with the one of our main specification equation (1). In contrast, for those centers opening and closing multiple times we firstly define each opening (closure) as an event and then assign housing units to each event according to a criterion based on time proximity.[16] As one can observe from Tables A5 and A6, our results are very similar following this alternative approach.

[15] See e.g. Goodman-Bacon (2021) explaining the source of the bias due to the comparison of already treated with newly treated units in difference-in-differences regressions.

[16] More precisely, we compute the time between two consecutive events related to a given center, and we assign all housing units observed within the first half of this spell to the first event and all units observed in the second half of the spell to the second event.

D. Effects of African Share: Nonparametric Estimation

In order to explore the effect of the share of Sub-Saharan Africans on rental prices in greater detail, we conduct a nonparametric estimation by dividing the center-time-specific African-share measure into five segments. We estimate the following equation:

$$\begin{aligned}\ln(Rent)_h = \ln(Rent)_{hcmt} = \hat{\alpha} + \beta_0(Active_{hct} \times Prox_{hc}) + \mu_1(Active_{hct} \times Quintile1_{hct}) + \\ \beta_1(Active_{hct} \times Prox_{hc} \times Quintile1_{hct}) + \mu_2(Active_{hct} \times Quintile2_{hct}) + \\ \beta_2(Active_{hct} \times Prox_{hc} \times Quintile2_{hct}) + \mu_3(Active_{hct} \times Quintile3_{hct}) + \\ \beta_3(Active_{hct} \times Prox_{hc} \times Quintile3_{hct}) + \mu_4(Active_{hct} \times Quintile4_{hct}) + \\ \beta_4(Active_{hct} \times Prox_{hc} \times Quintile4_{hct}) + \hat{\theta}_c Active_{hct} + \hat{\delta}_c Prox_{hc} + \\ + \mathbf{H}_h' \mathbf{\Gamma} + \hat{\lambda}_c + \hat{\gamma}_m + \hat{\tau}_{y[t]} + \hat{\varepsilon}_{hcmt}\end{aligned} \tag{4}$$

This equation is an extended version of the regression we estimate in column 3 of Table 1. In Table 1 we include interactions with the binary variable $African$, which takes the value of one if the Sub-Saharan African population share in the nearest center at time t is above the median value in the sample (which equals 24%). Here, we instead include a set of dummy variables to explore possible nonlinearities. The variables $Quintile\#_{hct}$ are dummies set equal to one if the closest center to a given rental posting contains a share of Sub-Saharan African individuals above the #th quintile of African shares observed in the sample.[17] The sample African shares are 14% at the 1st quintile, 21% at the 2nd quintile, 28% at the 3rd quintile and 39% at the 4th quintile. Figure A1, where we report our estimated coefficients $\beta_1 - \beta_4$ of equation 4, shows a statistically significant discontinuity around the 3rd quintile.

[17] Note that a given observation can correspond to several quintile dummies taking a value of one. If an observation is e.g. at the 45th percentile, it would result in the two dummy variables $Quintile1_{hct}$ and $Quintile2_{hct}$ both being set to one, while the $Quintile3_{hct}$ and $Quintile4_{hct}$ dummies would be set to zero.

Appendix: Tables

Table A1: Spatial-Differencing Estimates of the Effect of Local Disamenities on Housing Prices

Citation	Country (region)	Spatial scope	Disamenity type	Time period	Identification strategy	# events (obs.)	Dep. var.	Main elasticity (s.e.)
Asylum centers:								
Van Vuuren, Kjellander and Nilsson (2019)	Sweden (Gothenburg)	5 min walking	Asylum center building sites	2014-2017	DID (continuous distance)	3 (2,234)	ln(price)	-0.040 (0.017)
Daams, Proietti and Veneri (2019)	Netherlands	0.5-1 km	Asylum centers	2009-2017	DID	75 (346,262)	ln(price)	0.004 (0.014)
Hennig (2021)	Germany (Berlin)	0.1 km	Asylum centers	2010-2018	DID + event study	177 (208,640)	ln(rent)	-0.035 (0.015)
Dröes and Koster (2023)	Netherlands	2 km	Asylum centers	1990-2015	DID + event study + triple DID	51 (194,436)	ln(price)	-0.060 (0.009)
This paper	Switzerland	0.7 km	Asylum centers	2004-2014	DID + event study + triple DID	91 (154,708)	ln(rent)	-0.038 (0.012) [all centers] -0.026 (0.011) [< median African] -0.048 (0.015) [> median African]
Other disamentities:								
Pope (2008*a*)	USA (Wake County NC)	Noise level contours	Airport noise	1992-2000	DID	1 (16,900)	ln(price)	-0.031 (0.008)
Pope (2008*b*)	USA (Florida)	0.1 miles	Sex offenders	1996-2006	DID	322 (5,923)	ln(price)	-0.023 (0.012)
Linden and Rockoff (2008)	USA (N. Carolina)	0.1 miles	Sex offenders	1994-2004	DID	396 (169,557)	ln(price)	-0.041 (0.020)
Boes and Nüesch (2011)	Switzerland (Zurich)	Noise level	Airport noise	2001-2006	DID	1 (6,562)	ln(rent)	-0.047 (0.017)
Congdon-Hohman (2013)	USA (Ohio)	0.25 miles	Methamphetamine labs	2004-2007	DID + event study	101 (2,791)	ln(price)	-0.101 (0.049)
Currie, Davis, Greenstone and Walker (2015)	USA	0.5 miles	Toxic plants	1998-2005	DID + event study	2,171 (1,114,248)	ln(price)	-0.014 (0.007)
Muehlenbachs, Spiller and Timmins (2015)	USA (Pennsylvania)	1.5 km	Shale gas	1995-2012	DID-matching + triple DID	1 (1,835)	ln(price)	-0.100 (0.046)
Taylor, Phaneuf and Liu (2016)	USA (Minneapolis- Saint Paul)	0.7 miles	Hazardous waste sites	1990-2007	DID + event study	103 (95,194)	ln(price)	-0.077 (0.004)
Dealy, Horn and Berrens (2017)	USA (Linn County OR)	0.1 miles	Methamphetamine labs/decontamination	1999-2013	DID + event study	99 (8,025)	ln(price)	-0.027 (0.016)
Liang and Alexeev (2022)	Australia (Melbourne)	0.8 km	Injection centers	2016-2021	DID + triple DID	1 (5,847)	ln(price)	-0.060 (0.023)
Guignet, Jenkins, Nolte and Belke (2023)	USA	5.75 km	Chemical accidents	2004-2019	DID + event study	3,236 (10,427,165)	ln(price)	-0.031 (0.006)
Muñoz-Morales and Singh (2023)	USA	Attendance boundary	School shootings	1995-2017	DID + event study	11 (34,808)	ln(price)	-0.021 (0.007)
Tang and Le (2023)	USA (Baltimore)	0.1 miles	Gun offenders	2004-2019	DID + event study	1,411 (47,001)	ln(price)	-0.176 (0.023)

Notes: The table summarizes the main estimated elasticities in papers analyzing the influence of various disamenities on local housing markets through spatial differences-in-differences (DID). The elasticities reported here are those that, due to the generality of the sample and the employed estimation technique, closely align with the methodology of this paper. They may therefore not always be those that are highlighted by the authors themselves.

Table A2: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Panel A: Characteristics of rental postings					
Yearly Rent per sqm (CHF)	279.7	94.0	15.7	2620	154,708
Surface (sqm)	80.2	37.0	11.0	1125	154,708
Standard (single-floor apartment)	0.820	0.384	0	1	154,708
Duplex	0.063	0.242	0	1	154,708
Attic	0.031	0.172	0	1	154,708
Studio	0.014	0.118	0	1	154,708
Furnished apartment	0.037	0.190	0	1	154,708
Terrace-apartment	0.002	0.046	0	1	154,708
Independent house (villa)	0.023	0.150	0	1	154,708
Row house	0.006	0.076	0	1	154,708
Semi-detached house	0.003	0.052	0	1	154,708
Farm	0.001	0.033	0	1	154,708
Other type of housing unit	0	0.022	0	1	154,708
Less than 2 rooms	0.131	0.338	0	1	154,708
2-2.5 rooms	0.211	0.408	0	1	154,708
3-3.5 rooms	0.323	0.468	0	1	154,708
4-4.5 rooms	0.245	0.430	0	1	154,708
5 rooms or more	0.090	0.286	0	1	154,708
Panel B: Center-specific variables (by rental posting)					
Dummy for location within 709 m of center ($Prox$)	0.246	0.431	0	1	154,708
Dummy for closest center being open ($Active$)	0.704	0.456	0	1	154,708
Share of sub-Saharan Africans in closest center ($African$)	0.277	0.168	0	1	154,708
Estimated crime propensity of population in closest center ($Crime$)	0.003	0.002	0	0.021	154,708
Average genetic distance of center population w.r.t. native Swiss population ($Gendist$)	0.019	0.007	0	0.051	154,708
Average local genetic distance between local and center populations ($Local\ Gendist$)	0.013	0.004	0.003	0.027	153,578
Share of Muslim asylum seekers in closest center ($Muslim$)	0.557	0.156	0	1	154,682
Average education level of local population ($Education$)	0.163	0.047	0.071	0.289	154,708
Panel C: Center characteristics					
Hosting capacity	95	88	30	694	91
Duration of opening (years)	14	8	4	50	91

Notes: The table reports summary statistics for our sample of rental postings and centers. Characteristics of rental postings were obtained from the consulting firm meta-sys.ch. Information on asylum seekers ($African$, $Crime$, $Muslim$) was obtained from Couttenier et al. (2019), data on genetic distances ($Gendist$ and $Local\ Gendist$) are from Spolaore and Wacziarg (2018), and education are from the Federal Statistical Office. The descriptive statistics shown here are for the continuous version of those variables. In our estimations, we transform them into binary variables, using the median value as cutoff unless otherwise stated. The data on hosting center capacity and opening duration were obtained from 13 cantonal authorities or private bodies mandated by the cantons to manage the centers. For the list of included cantons, see Table A3.

Table A3: Number of Centers and Rental Postings per Canton

Canton	Centers	Rental postings
Aargau	21	22,405
Geneva	17	8,778
Glarus	1	603
Graubünden	4	9,053
Jura	2	501
Neuchâtel	2	581
Schaffhausen	2	3,441
Solothurn	3	576
Thurgau	2	5,895
Ticino	1	1,609
Valais	7	1,978
Vaud	8	25,514
Zurich	21	73,774
Total	91	154,708

Notes: The table reports the total number of centers and rental postings observed in the baseline estimation sample for each canton.

Table A4: Validation of Crime Propensity Measure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable	Crime AS	Crime All	Crime AS	Crime All	Crime AS	Crime All	Crime AS	Crime All
$Active$	-1.956	-1.950	-2.467	-1.864	-1.718	-1.832	-2.231	-1.753
	(2.311)	(2.101)	(2.221)	(1.897)	(2.309)	(2.091)	(2.209)	(1.900)
$Active \times \ln(Capacity)$	0.208	0.406	0.234	0.402	0.226	0.406	0.252	0.403
	(0.502)	(0.498)	(0.505)	(0.485)	(0.500)	(0.499)	(0.503)	(0.485)
$Active \times Crime$			0.521**	-0.095			0.520**	-0.092
			(0.251)	(0.228)			(0.251)	(0.228)
$Active \times African$					-0.380	-0.180	-0.376	-0.173
					(0.431)	(0.256)	(0.405)	(0.259)
Observations	4,184	4,184	4,184	4,184	4,184	4,184	4,184	4,184

Notes: For the estimations reported in this table, the unit of observation is a municipality-month, ms. The sample consists of the municipalities for which we have information on at least one hosting center. The dependent variable is the number of violent or property crimes committed by asylum seekers (full population) in odd (even) columns. In each column we control for the natural logarithm of the population living in the municipality at yearly frequency as well as for municipality, year and calendar month fixed effects. Clustered standard errors by municipality are reported in parentheses. The total number of municipalities (clusters) present in the sample is 59. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Stacked DiD: Baseline Estimates

	(1)	(2)	(3)	(4)
Dep. variable	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)
Radii in meters:	*baseline*			
treatment; spillover	709; 1,110	709; 1,425	500; 1,000	500; 1,500
$Active \times Prox$	-0.041***	-0.032***	-0.057***	-0.048***
	(0.012)	(0.011)	(0.016)	(0.016)
Observations	154,707	106,833	137,010	78,277
R^2	0.453	0.446	0.456	0.450

Notes: For the estimations reported in this table, the unit of observation is a rental posting, $h = hcmt$. The sample covers rental postings for housing units located within 2km of a hosting center, for the period 2004-2014. In each column we control for a set of housing characteristics (described in Section 1.2.1), year and municipality fixed effects, as well as the fixed effects to exploit within event variation $\tilde{\theta}_e Active_{het}$, $\tilde{\delta}_e Prox_{he}$ and $\tilde{\lambda}_e$ in accordance with equation 3. The fixed-effects structure we impose compares rental postings within each event of center opening (closing) (stacked DiD, see Appendix C for further details). Clustered standard errors by municipality are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Stacked DiD: Socio-Ethnic Differences in Center Populations

	(1)	(2)	(3)	(4)
Dep. Variable	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)
Effect	Base	Crime	African	African & Crime
$Active \times Prox$	-0.0409***	-0.0335***	-0.0282***	-0.0254**
	(0.0116)	(0.0121)	(0.0101)	(0.0102)
$Active \times Crime$		0.0060		0.0057
		(0.0046)		(0.0052)
$Active \times Prox \times Crime$		-0.0115**		-0.0061
		(0.0058)		(0.0059)
$Active \times African$			0.0006	-0.0006
			(0.0087)	(0.0087)
$Active \times Prox \times African$			-0.0247***	-0.0227***
			(0.0050)	(0.0059)
Observations	154,707	154,707	154,707	154,707
R^2	0.453	0.453	0.453	0.453

Notes: For the estimations reported in this table, the unit of observation is a rental posting, $h = hcmt$. The sample covers rental postings for housing units located within 2km of a hosting center, for the period 2004-2014. In each column we control for a set of housing characteristics (described in Section 1.2.1), year and municipality fixed effects, as well as the fixed effects to exploit within event variation $\tilde{\theta}_e Active_{het}$, $\tilde{\delta}_e Prox_{he}$ and $\tilde{\lambda_e}$ in accordance with equation 3. The fixed-effects structure we impose compares rental postings within each event of center opening (closing) (stacked DiD, see Appendix C for further details). Clustered standard errors by municipality are reported in parentheses. The sample number of municipalities (clusters) is 192. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Alternative Dimensions of Heterogeneity: Muslim and Genetic Distance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. Variable	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)	ln(Rent)
Effect	Crime	Muslim	Gendist	African	Gendist-Crime	Gendist-Muslim	African-Crime	African-Muslim	Gendist-Crime-Muslim	African-Crime-Muslim
$Active \times Prox$	-0.0314**	-0.0378***	-0.0279***	-0.0266**	-0.0254**	-0.0259**	-0.0239**	-0.0246**	-0.0233**	-0.0217**
	(0.0129)	(0.0118)	(0.0107)	(0.0107)	(0.0110)	(0.0108)	(0.0108)	(0.0106)	(0.0109)	(0.0106)
$Active \times Crime$	0.0067				0.0068		0.0065		0.0071	0.0067
	(0.0045)				(0.0049)		(0.0051)		(0.0049)	(0.0052)
$Active \times Prox \times Crime$	-0.0107*				-0.0053		-0.0058		-0.0056	-0.0061
	(0.0057)				(0.0059)		(0.0059)		(0.0057)	(0.0058)
$Active \times Muslim$		-0.0023				-0.0030		-0.0024	-0.0033	-0.0028
		(0.0055)				(0.0049)		(0.0049)	(0.0048)	(0.0048)
$Active \times Prox \times Muslim$		-0.0011				-0.0039		-0.0041	-0.0038	-0.0040
		(0.0087)				(0.0089)		(0.0084)	(0.0085)	(0.0081)
$Active \times Gendist$			-0.0024		-0.0038	-0.0029			-0.0044	
			(0.0081)		(0.0077)	(0.0078)			(0.0075)	
$Active \times Prox \times Gendist$			-0.0201***		-0.0184***	-0.0209***			-0.0190***	
			(0.0053)		(0.0064)	(0.0061)			(0.0072)	
$Active \times African$				0.0008			-0.0006	0.0003		-0.0012
				(0.0084)			(0.0083)	(0.0081)		(0.0080)
$Active \times Prox \times African$				-0.0221***			-0.0203***	-0.0229***		-0.0209***
				(0.0049)			(0.0059)	(0.0053)		(0.0063)
Observations	154,708	154,682	154,708	154,708	154,708	154,682	154,708	154,682	154,682	154,682
R^2	0.453	0.453	0.453	0.453	0.453	0.453	0.453	0.453	0.453	0.453

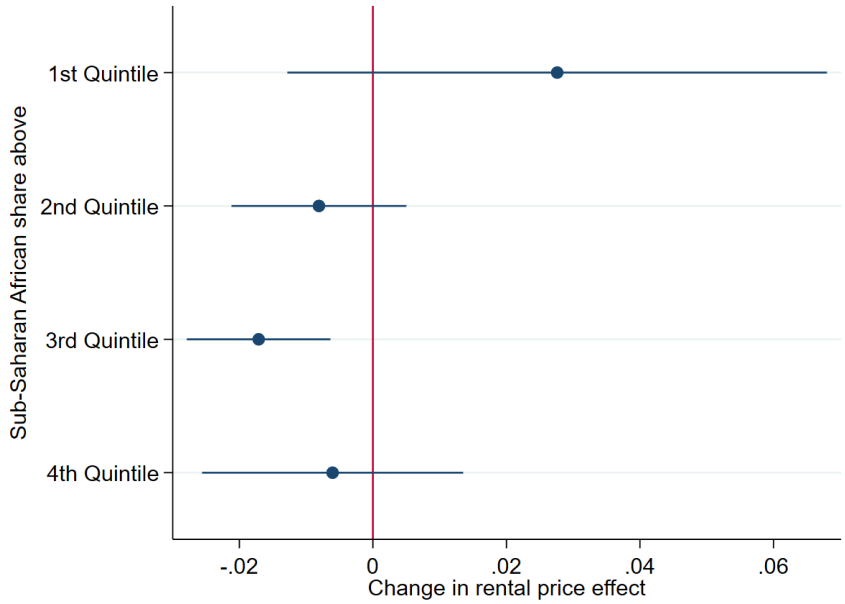
Notes: For the estimations reported in this table, the unit of observation is a rental posting, $h = hcmt$. The sample covers rental postings for housing units located within 2km of a hosting center, for the period 2004-2014. In each column we control for a set of housing characteristics (described in Section 1.2.1), year and municipality fixed effects, as well as the fixed effects to exploit within center variation $\theta_c Active_{hct}$, $\delta_c Prox_{hc}$ and λ_c in accordance with equation (1). Clustered standard errors by municipality are reported in parentheses. The sample number of municipalities (clusters) is 192. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Alternative Cutoffs for Binary Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Ln Rent	Ln Rent	Ln Rent	Ln Rent	Ln Rent	Ln Rent
Effect	Perc 50	Perc 60	Perc 75	Perc 50	Perc 60	Perc 75
$Active \times Prox$	-0.0266**	-0.0302***	-0.0313***	-0.0239**	-0.0283***	-0.0309***
	(0.0107)	(0.0105)	(0.0119)	(0.0108)	(0.0107)	(0.0113)
$Active \times African50$	0.0008			-0.0006		
	(0.0084)			(0.0083)		
$Active \times Prox \times African50$	-0.0221***			-0.0203***		
	(0.0049)			(0.0059)		
$Active \times African60$		0.0079			0.0065	
		(0.0070)			(0.0070)	
$Active \times Prox \times African60$		-0.0188***			-0.0178**	
		(0.0054)			(0.0072)	
$Active \times African75$			0.0212			0.0231
			(0.0135)			(0.0146)
$Active \times Prox \times African75$			-0.0214**			-0.0204*
			(0.0103)			(0.0119)
$Active \times Crime50$				0.0065		
				(0.0051)		
$Active \times Prox \times Crime50$				-0.0058		
				(0.0059)		
$Active \times Crime60$					0.0073	
					(0.0063)	
$Active \times Prox \times Crime60$					-0.0039	
					(0.0067)	
$Active \times Crime75$						-0.0096
						(0.0063)
$Active \times Prox \times Crime75$						-0.0015
						(0.0075)
Observations	154,708	154,708	154,708	154,708	154,708	154,708
R^2	0.453	0.453	0.453	0.453	0.453	0.453

Notes: This table reports estimates for different variants of the binary variables $African$ and $Crime$, where the suffix '50' denotes the sample median as the cutoff value, '60' denotes the 60th percentile, etc. These estimations serve as robustness tests of the results shown in columns (3) and (4) of Table 1. For the estimations reported in this table, the unit of observation is a rental posting, $h = hcmt$. The sample covers rental postings for housing units located within 2km of a hosting center, for the period 2004-2014. In each column we control for a set of housing characteristics (described in Section 1.2.1), year and municipality fixed effects, as well as including fixed effects to exploit within-center variation $\theta_c Active_{hct}$, $\delta_c Prox_{hc}$ and λ_c in accordance with equation (1). Clustered standard errors by municipality are reported in parentheses. The sample number of municipalities (clusters) is 192. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A1: Effects of African Share: Nonparametric Estimation



Notes: The figure shows triple interaction estimates from equation 4. For example, the coefficient for ' 1st Quintile' the marginal change in rental price effect when the Sub-Saharan African share of the nearest center rises above the first quintile of Sub-Saharan African shares observed in the sample. The baseline (omitted) category contains rental postings with Sub-Saharan African shares of their nearest hosting centers below the 20th percentile.