The Logic of State Surveillance:

Evidence from a Century of Political Dissent

Gemma Dipoppa * Annalisa Pezone †

March 3, 2025

Abstract

All states adopt systems to surveil political activists. How do they decide whom to watch and why? We explore the strategic calculus behind state surveillance with the universe of Italian files (N=150,000 individuals) over a century (1816–1945), spanning democracy and autocracy. We argue that states perceive their greatest threat to stability as individuals who combine mobilization potential (through education) and grievances (arising from marginalization). Leveraging a discontinuous expansion in primary schooling, we document an increase in surveillance targeting more educated cohorts, driven by intensified monitoring of the working class. The education shock did not affect political mobilization, suggesting that state fears — rather than individual activism — drove the increase in surveillance. These findings underscore how socioeconomic resources, jointly linked to political activation, are decoupled in the minds of the state. States perceive education as a tool for political empowerment, but view marginalization as a source of radicalization, making educated members of marginal groups primary targets of surveillance.

*Columbia University, Department of Political Science, 420 W. 118th Street, NY 10027 (gd2733@columbia.edu)

†New York University, Department of Politics, 19 W. 4th Street New York, NY 10012 (annalisa.pezone@nyu.edu).

1 Introduction

States worldwide use systems of surveillance to monitor the political activity of their citizens. While surveillance is necessary to maintain security, it may also infringe on individual rights and be used to suppress democratic movements (Andirin et al., 2022). Motivated by the proliferation of increasingly precise systems for controlling citizens, and political activists in particular, a recent literature examined how state surveillance in autocratic regimes affects dissent (Xu, 2021; Hager and Krakowski, 2022; Beraja et al., 2023) and interpersonal trust (Lichter, Löffler and Siegloch, 2021).

However, the logic behind state surveillance and the factors determining who states decide to watch remain largely unexplored. Gaining insight into these choices is critical to understanding political change, as the success of political movements depends not only on participation, but also on the barriers that may inhibit it. Yet, partly due to security concerns shielding states from being accountable for their intelligence choices, we know surprisingly little about how states strategically invest their limited monitoring resources.

In this paper, we study states' strategies of surveillance using a newly collected database of the universe (N=152,000) of individuals under political monitoring in Italy born between 1816 and 1932, and surveilled from 1869 to 1946. The data spans democratic and autocratic regimes, encompasses a range of political movements (from anarchism and communism to fascism), and provides detailed activists' information, offering a unique opportunity to study state's choices across contexts, including – unlike in previous work – in democratic settings.

The theoretical point of departure of this paper is that while individuals with socioeconomic resources – high income and education – are more likely to engage in political action (Almond, Verba et al., 1963; Verba et al., 1993), individuals with low income, or other types of marginal status, might be perceived by states as the most dangerous – the most likely to engage in revolutionary movements aimed at overthrowing regimes (Marx and Engels, 1848; Gramsci, 1975; Fanon, 1963; Gurr, 1970; Hobsbawm, 1971; Skocpol, 1979; Scott, 1985). Historically, from the French to the Russian, from the Haitian to the Salvadoran revolutions, uprisings

that led to regime changes have seen economically and socially marginalized groups as the political agents uprooting existing structures. At a micro-level, qualitative evidence from the reports of the Italian Surveillance System indicates that the state watched individuals embedded in low socioeconomic environments but with higher education than their peers. While education of all groups is likely to trigger increased participation, that of low-income, or otherwise marginal, individuals might be more likely to lead to revolutionary actions and thus be considered by the state as particularly concerning for stability. We thus propose the hypothesis that states prioritize surveillance of individuals with high education but limited economic resources, as they perceive them as the most dangerous for regime survival.

Offering evidence for this hypothesis is challenging for at least two reasons: First, it requires observing whom the state watches, including information on the characteristics of those surveilled. Second, individual traits like education and income are endogenous and correlated with other factors influencing surveillance, making it difficult to establish a causal link. As a result of these complexities, while seminal work exists on the determinants of surveillance – finding support for racial profiling of black neighborhoods in Detroit (Davenport, 2005) – it remains challenging to isolate the factors prompting state tracking.

To tackle the data challenge, we use declassified data on activists surveilled by the Italian state which includes detailed biographical information. We address the endogeneity challenge by taking advantage of an exogenous shock to education predominantly affecting middle and lower classes: an Italian law increasing public schooling at a time when high-income pupils largely attended private institutions. This law applied discontinuously in time, extending primary education by two additional years for cohorts born after its approval, and in space, to cities above a 4,000 inhabitants population cutoff.[1] We provide evidence that this education reform was effective: cross-sectional regression discontinuity results indicate that municipalities affected by the law displayed lower illiteracy rates several years after the reform compared to municipalities just below the threshold. This aligns with historical

[1]No other law applied discontinuously to this population threshold and cohorts.

accounts emphasizing the reforms' role in reducing illiteracy among the lowest classes. Our empirical strategy tests the impact of the reform on surveillance in marginally treated municipalities-cohorts: we compare cohorts born right before and after the law would impact them, located in municipalities right below and above the population cutoff in a differences-in-discontinuities (DDRD) framework (Grembi, Nannicini and Troiano, 2016).

In municipalities-cohorts affected by the education reform, the share of individuals surveilled increases by 3.5 per 100,000, a substantial increase from the baseline of 0.08. This effect is absent in pre-reform cohorts, consistent with the parallel-trends assumption. Notably, when a later reform removes the 4,000-inhabitant education discontinuity and standardizes the law across municipalities, the effect disappears, reinforcing that the schooling shock – not other factors – drives the increase in surveillance. The result holds across specifications, including a stricter "double" DDRD controlling for both population and distance to the first treated cohort's birth date, and a more flexible difference-in-differences approach. The effect remains consistent across alternative models and dependent variable choices, and is absent at placebo cutoffs and cohorts.

Was the increase in surveillance driven by marginal groups? We use information about activists' professions to categorize them into economic classes. The increase in surveillance is predominantly driven by working class individuals. Given that the reform likely reduced the proportion of the working-class population, this outcome may represent a lower bound of the reform's impact on surveillance of the working class.

The finding that education increases surveillance could stem from two forces: state concern with the political activities of educated individuals in lower socioeconomic classes, and their increased political mobilization due to higher access to education. Evidence indicates that the second mechanism plays a limited role in this setting. We leverage partly newly collected data on five measures of political activism ranging from low to high cost of action: (1) voter turnout in national elections, (2) records of notable individuals engaged in politics which we source from Wikidata, (3) data on protests and strikes that we scrape

from historical newspaper archives, (4) the universe of partisans scraped from an online repository, and (5) instances of banditry documented by Lecce, Ogliari and Orlando (2022). Across measures and specifications, our findings show null effects of the education reform on political activation. Although the lack of effects on individual activation could be the result of increased surveillance, we did not find significant differences in political engagement in response to reform between areas with high and low surveillance rates or between cohorts that were never under surveillance. Regardless of surveillance, state-provided education did not increase average political engagement or action against the state.

If not individual but, rather, state activation is the underlying mechanism, we should find evidence in line with two observable implications: that marginalized but educated individuals, which the state considered a threat, are subject to more intense surveillance, and that surveillance is targeted at ideologies advocating for marginalized groups. Our findings support these expectations. Individuals in treated municipalities-cohorts are (i) more frequently assigned to political confinement, a costly surveillance tool that monitors and sometimes forcibly relocates dissidents, (ii) face prolonged surveillance, and (iii) the volume of information collected on treated cohorts is higher. For each of these findings, the intensity of surveillance is highest for working class individuals. We also find that communists and socialists – catering to the demands and aspirations of the working class – are the most surveilled ideologies. Finally, we show that, regardless of regime type, whether democracy or autocracy, and of state capacity, states consistently monitor politically engaged, marginalized groups whom they perceive as potential threats to stability.

In the final part of the paper, we provide support against alternative explanations for our findings. First, we show that education-driven migration to larger cities, where activism might be higher, does not drive our findings by replicating the results on non-migrants. Second, we test whether increased surveillance could be attributed to an increase in the number of state officials acting as informants. While data on informants are absent, we leverage a subsequent law discontinuously increasing the number of state officials without

affecting education, and find no evidence that additional bureaucrats led to increased surveillance. We also do not find that teachers are surveilled at higher rates post-reform, ruling out that state concern for their activity as instigators of political action drives the findings. Lastly, since the state often replaced the church in providing education, it is likely that this substitution would, if anything, reduce anti-state mobilization and the need for surveillance, since states use education to foster compliance.

This study falls at the intersection of several strands of research. An important body of work across social sciences has investigated the multiple ways in which states repress political dissent. While several studies have focused on overt forms of state repression, such as arrests and killings,[2], a much smaller and recent literature has studied covert repression – surveillance. Prior work on this topic, with the exception of the seminal study by Davenport (2005), has focused on the consequences of surveillance on political engagement (Andirin et al., 2022; Hager and Krakowski, 2022; Xu, 2021; Beraja et al., 2023) and trust (Lichter, Löffler and Siegloch, 2021). In contrast, this paper considers surveillance as a dependent variable and studies the strategic determinants of citizens' surveillance by the state. While Davenport (2005) examined the correlates of surveillance in a single movement in Detroit, our analysis spans across regimes and ideologies, and identifies the impact of single individual characteristics leveraging an exogenous shock. In the conclusions section, we discuss why we believe that our findings based on historical data are still relevant for understanding a time of digital surveillance.

This paper also relates to a foundational areas of inquiry for social science, studying who participates in politics (Almond, Verba et al., 1963) and the conditions under which revolutions can take place (Kuran, 1989; Acemoglu and Robinson, 2001; Bueno de Mesquita

[2]Examples include Pierskalla (2010); Shadmehr (2014); Ritter and Conrad (2016); Klor, Saiegh and Satyanath (2016); Sullivan (2016); Blaydes (2018); Garfias and Magaloni (2018); Rozenas and Zhukov (2019); Truex (2019); Young (2019); Nugent (2020); Pan and Siegel (2020); Thachil (2020); Esberg (2021).

and Smith, 2017). Studies have identified marginalization – in terms of class (Marx and Engels, 1848; Gramsci, 1975), income (Scott, 1985), race (Fanon, 1963), and other factors (Hobsbawm, 1971; Skocpol, 1979) – as a key predictor of revolutions. We build on this tradition of inquiry to argue that states are particularly concerned with individuals who have high levels of education (mobilization potential) and low income (perceived propensity for revolt). By linking education to surveillance, our study also speaks to research on education and state control: While studies show that state-led education fosters obedience on average (Paglayan, 2024), our findings indicate it can also spark state fears of unrest. In this sense, we contribute to painting a more nuanced picture of states' trade-offs in offering schooling (Bourguignon and Verdier, 2000).

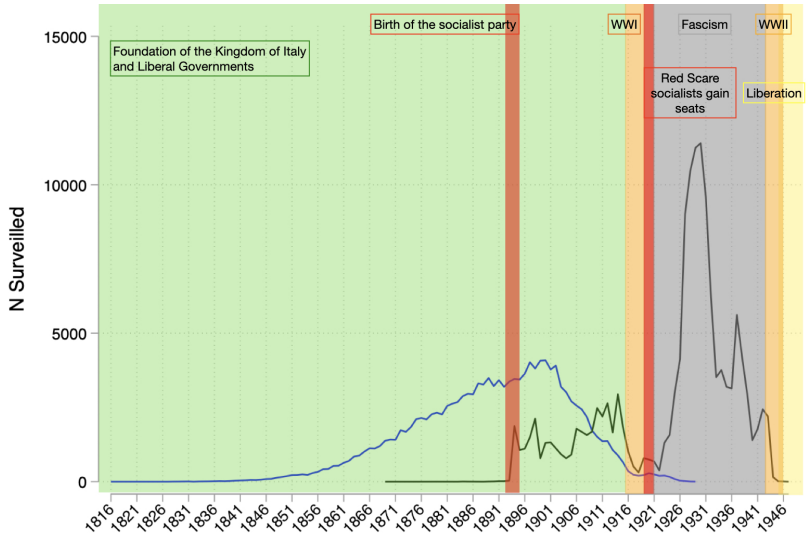
Our findings have significant implications for the understanding of state behavior and political change. They reveal a striking pattern: high socioeconomic status can simultaneously increase political engagement and protect individuals from state scrutiny. Instead, as marginalized groups gain education and political voice, they also face disproportionate surveillance compared to their equally mobilized but higher-status counterparts. This asymmetry can undermine political agency across groups, potentially muting the demands of society's most vulnerable members precisely when their civic participation increases.

2 State Surveillance

2.1 The Evolution of Surveillance in Italy

State surveillance in Italy has its roots in the systems established by the Kingdom of the Two Sicilies and Piedmont-Sardinia. The Bourbon monarchy in the Two Sicilies employed a centralized political police system inspired by the Napoleonic model, which was expanded after the 1820-21 and 1848 revolutions to include spies, informants, and surveillance of public spaces like theaters and squares to monitor those plotting against the monarchy (Di Fiore, 2018). In the same period, the Kingdom of Piedmont-Sardinia introduced the La Marmora

Figure 1: Surveilled by date of birth (blue) and of surveillance (black)



reform (Law n. 1676, 1854), creating the first secret service focused on military threats and political unrest.

Following Italy's unification in 1861, the new constitutional monarchy inherited and combined surveillance practices from the Bourbon and Piedmontese regimes. Over time, military intelligence expanded, and in 1884, a naval intelligence service was formed. As socialist and anarchist movements gained influence, the state centralized surveillance by creating a national database in 1894. While this is the official start of our data, the central office effectively consolidated information previously held by local police units and which had been collected over a long time span (Tosatti, 1990, p. 447-449). While the outbreak of World War I reduced police availability, the period saw innovations in surveillance technology, with Italy's Legione Sacra pioneering the use of telephonic wiretapping.

During Mussolini's regime, surveillance became more centralized and efficient. Inspired by Soviet model, the Äeka, Mussolini created the OVRA (Organization for the Vigilance and Repression of Anti-Fascism) in 1926, expanding the use of informants and spies to monitor political dissidents and intellectuals (Leto, 1951, p.46). The number of individuals surveilled surged to an average of 4,800 per year, up from 1,360. Operating with extensive powers,

the OVRA played a pivotal role in maintaining Fascist control, changing from a tool of monarchical surveillance into an instrument of totalitarian control.

2.2 The Architecture of State Surveillance

Qualitative archival work, such as Dilemmi (2010)'s study of Verona's surveillance archives, reveals how individuals came under state scrutiny. Most often, state attention was triggered by participation in political parties, unions, or opposition movements, involvement in protests or strikes, subscribing to critical publications, or associating with others already under surveillance (Dilemmi, 2010, p. 215).

A significant portion of the files in the Verona archives originated from the Police Headquarter (43%), which generally began with reports by the local police. In other instances, reports were started by higher institutional levels, such as the central Ministry of Interior (18%) and the provincial Carabinieri (18%) (Dilemmi, 2010, p.206). Information on political activists could also occasionally come from informants: in some cases, those were individuals with jail sentences who would report on their cellmates in exchange for a shortening of the sentence (Storia Criminale, Ciconte, Ch. 9). With the raise of fascism, new actors, including a proliferation of fascist regime officials and organizations, increasingly contributed to initiating surveillance cases (Dilemmi, 2010, p.258).

Individuals under state monitoring were generally unaware of being surveilled. However, a considerable fraction of those surveilled were also targets of additional special measures, and some of those were aware of being surveilled. These measures shifting surveillance from cover to overt, in increasing order of severity, included the *diffida* (a formal warning to maintain good conduct, 5.5% of the sample), registration in the *Rubrica di frontiera* (surveillance of an individual's movement across national borders, sometimes discreetly, 23%), and *confino* (forced relocation, 30%).

In some cases, surveillance files were closed when the person under observation was no longer deemed a threat. This could occur after a period of inactivity, when the individual

ceased their involvement in activities that initially raised suspicion, or when political and social circumstances changed, leading authorities to reconsider their priorities. The decision to close a file typically reflected a belief that the individual no longer posed a danger to public order or to the regime's stability.

2.3 Targets of Monitoring: Qualitative Evidence

Who should states be afraid of? A long tradition of thought across social sciences has laid the theoretical foundations for thinking about the conditions under which revolutions can take place. While some have considered what happens when leaders appear weak (Bueno de Mesquita and Smith, 2017), others have focused on who rebels, emphasizing that this is connected to political participation and, more specifically, by the capacity to mobilize for collective action (Kuran, 1989; Acemoglu and Robinson, 2001).

However, not all forms of mobilization are equally threatening. From the state's perspective, individuals who are both marginalized and newly equipped with political awareness—whether due to class (Marx and Engels, 1848; Gramsci, 1975), race (Fanon, 1963), or other social characteristics (Hobsbawm, 1971) —pose a unique challenge. Gurr (1970) concept of relative deprivation captures this dynamic, suggesting that rebellion is most likely when individuals not only experience exclusion but also become aware of it.

Under these conditions, education can function as a catalyst, transforming latent grievances into an organized political challenge. Thus, when the state perceives this conjunction of education and marginalization as a threat, it may respond with preemptive measures to neutralize the perceived risk before it materializes into mobilization.

Our hypothesis is that individual characteristics –education and income– play a role in determining the likelihood of being surveilled. A qualitative indication that states might consider these factors when deciding whom to monitor can be drawn from an analysis of the files written by the political police in Italy. While there is no systematic digital collection of the full write-ups, we rely on the work of Ciancarini (2019) and Carolini and Ecca (2015),

who transcribed the files of 254 dissidents from Civitavecchia and 1,000 antifascists from 1927 to 1943, respectively. Neither source constitutes a random sample –the specific location and ideology of these individuals may make these files different from others–, but the biographical information contained within them offers valuable insight into what aspects of a person's life the state deemed important.

In these summaries, references to the education or intelligence of the individuals under surveillance[3] appear in 21% of the observations, frequently accompanied by an evaluation of the individual's capacity to influence public opinion. Some examples are reported below:

> "His listed profession is mechanical blacksmith [...] He enjoys a good reputation in the public. He has a bold and enterprising character. His education is above his status. Quick intelligence. Limited culture but above his status. He completed primary education." (CPC 75)

> "By profession a stonemason [...] the police, in their reports, acknowledge him to have a good reputation in public opinion, a rather mild character, good manners, sharp intelligence." (CPC 33)

> "Mechanic (although the police record him as a dock worker). The Prefecture, on March 31, 1930, judged him to enjoy a good reputation, of mild character, good manners, fairly intelligent, of decent education." (CPC 1119)

Education and access to knowledge were sufficiently important to the state that, when possible, deliberate restrictions were taken to suppress its use. It was the case for the antifascist and politician Riccardo Bauer, imprisoned during Fascism:

> "He served his sentence in the penal house of Alessandria. In 1932, he is denied permission to read and write during detention." (AANPIA)

[3]We consider both factors together, as they are often mentioned jointly and convey similar information to the police.

Together with education, the state also collected information on the profession of the surveilled. As evidenced in the quotes and described in the data section, profession was recorded for 100% of the surveilled, highlighting how this aspect of individual identity was fundamental for the state, as important as documenting the political stance of the dissidents –another key piece of information systematically noted for each subject.

3 Data

We rely on a broad range of historical, census, and administrative data. Below, we outline the key regressors and outcomes, while additional details are provided in Section A in the Appendix.

State Surveillance: Our primary dependent variable of interests is state surveillance of political activists. We scrape individual level data on 152,434 individuals surveilled by the Italian State from the *Italian Central Political Surveillance Office* (Casellario Politico Centrale). For each person, the archive contains name, date of birth, location of birth and residence, political ideology, occupation, and the years of surveillance. We include an example of a biographical file in Figure B.6. After linking with the first population Census of 1861 and excluding individuals born outside of Italy, we obtain a sample of 102,535 individuals born from 1816 to 1929, and surveilled from 1869 to 1947. While the national archive was officially established in 1894, the police started collecting information on political activists before this date and included these files in the archive once it was centralized. Figures B.1 and B.2 present the spatial and temporal variation of the number of individuals surveilled in 100,000 inhabitants during the period under study. At least one individual was surveilled in 78% of the municipalities, with a larger concentration of surveilled in Central Italy, an area that saw intense political activism from the Risorgimento to the years of the Socialist Movement. Consistent with progressive improvements in state capacity, the number of surveilled increases throughout the period.

The information contained in the police archive allows us to categorize individuals into upper, middle, and working class. As social classifications may change over time, we refer to Labini (1972), whose work provides a framework for classifying occupations in 1881 into classes, reflecting the societal norms of that era. Figures B.3 and B.4 report the distribution of occupations across the different classes, highlighting the most common jobs and the absolute number of individuals per group. The majority of surveilled comes from the working class (67.21%) –unskilled laborers, farmers, masons, and carpenters are the most common professions. The working class constituted 58% of the population, according to 1881 estimates (Labini, 1972), making this class slightly over-represented in the surveillance files, compared to its share in the population.

We use information in the files to code other characteristics of the surveilled. Figure B.4 considers ideology: about half of those surveilled were communists and socialists (52.41%), followed by anti-Fascists (22.94%), anarchists (17.84%), and other ideologies (6.81%). In Figure B.5 we consider gender: 2.82% of the surveilled were female, most of whom were housewives (27.61%), and tailors (4.58%).

Measures of Political Engagement: We collect five indicators of political activation, spanning from the least to the most resource and risk-intensive forms of mobilization. First, we obtain data on the number of votes casted in each municipality from the *Electoral Historical Atlas* archive. Since data on the voter base is unavailable, and since this base expanded over time –in ways likely endogenous to the variables studied in this paper– our analyses consider directly the absolute number of votes controlling for population at baseline.

Second, we collect information on notable individuals and we classify individuals known for their role in politics, broadly defined. We scrape data on all notable individuals born in Italy from 1816 to 1895 from WikiData, including information on their birth date and place, occupation, and other significant roles held. Using the last two entries, we identify notable individuals who were involved in politics, ranging from members of parliament to activists

and high level religious figures.

We move towards costlier forms of political action by scraping news items on the number of protests and strikes from the archives of one of the main Italian newspapers, *La Stampa*, the only including articles dating as far back as 1867. We identify the location of the event by writing an algorithm that searches the municipality name in the text of the article.

Our last two sources directly consider outright rebellion: we scrape a new database of the universe of partisans born between 1816 and 1895 (N=19,341), anti-fascist resistance fighters who engaged in armed struggle against the state. This database reunites regional sources that, after World War II, collected the lists of individuals officially part of the partisan movement. The data includes names, date and location of birth of the partisans, information which allows us to consider the same DDRD strategy as in the main specification.[4]

Finally, we obtained data on another highly costly form of political action: banditism. Bandits, or "briganti", were individuals who lived outside of the law and engaged in criminal acts such as raids, kidnappings, and thefts, particularly targeting the wealthy and landowners. Their actions were often a form of insurgency against the newly unified Italian state. Banditry was especially prevalent in southern Italy, and in parts of central Italy, where social inequalities were severe and many viewed bandits as rebels rather than simple criminals (Ciconte, 2011). We use data collected by Lecce, Ogliari and Orlando (2022) on the municipal level number of bandits in Southern Italy (the data excludes Sicily) from 1860 to 1870. For both protests and banditry data we lack information from the period before the education reform. This data is thus used in cross-sectional analyses. Additional details on each data source are discussed in Section A, SI.

[4]It is not possible, instead, to match partisans and surveilled: partisans started their activity in the summer of 1943, after Mussolini was arrested, a temporary government was put in place in Rome, while Northern Italy was under German occupation. Files of partisans might be present in the German archives, but the Italian secret police largely halted its surveillance activity in this period, as is visible in Figure 1.

Treatment and other covariates: Our treatment is defined based on population data, which we gather from the Historical Censuses made available by the Italian Institute of Statistic. Our year of reference is 1861, the first Census in Italian history.[5]. We also collect and digitize data on the share of illiterates at the municipal level from the 1911 Census, the first including municipal level data on education.

4 Empirical Framework

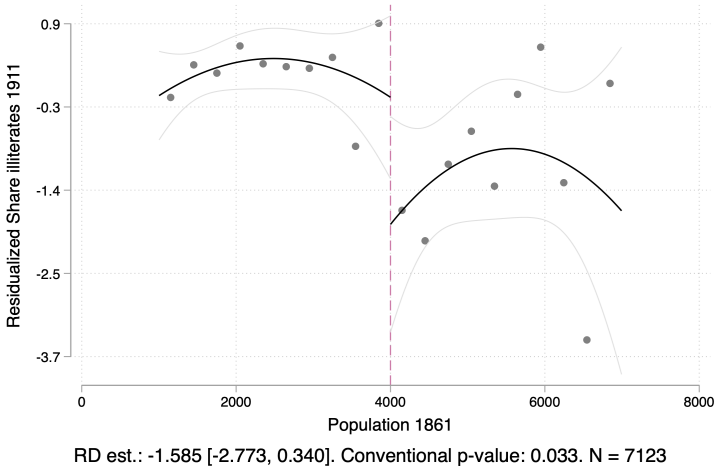
4.1 Treatment: The Education Reform

Our treatment of interest is a shock in the number of years of education resulting from the first education reform of the Italian Kingdom. The Casati Law, enacted across Unified Italy in 1861,[6] reformed the education system by introducing free elementary schooling for children aged six and older, and making the first two years of elementary school compulsory. Although the law applied universally, it required towns with more than 4,000 inhabitants to extend compulsory schooling by an additional two years, covering ages eight to ten (n. 3725/1859, art. 319-321). Consequently, children who were seven or younger at the time of the law's enactment and living in towns above the population threshold experienced a positive education shock compared to their peers born just before or in smaller municipalities. This discontinuity in education disappeared in 1904, when the Orlando Law extended compulsory schooling to age twelve for all municipalities, meaning that individuals born in or after 1896 received the same education irrespective of where they lived. Therefore, our analysis focuses on cohorts born before 1896.

[5]Municipalities in province of Udine started being part of Italy only in 1871, after the start of our treatment, and are therefore excluded from the sample.

[6]The law was initially approved in 1859 in the Kingdom of Sardinia, but its implementation was delayed due to the unification war. Before this time, no laws requiring compulsory education had been implemented (Sotgiu, 1984).

Figure 2: (First Stage) Education Reform Impact on Illiteracy (RDD)



	(1) Share Illiterates
RD_Estimate	-1.585** (0.744)
Observations	7,123
Bandwidth	2567.148
N Left	5916
N Right	1207
Robust P-val	0.125
Conventional P-val	0.033
Mean DV	33.21

Notes: The figure and table report the results from a cross-sectional regression discontinuity considering the impact of the education reform (applying to municipalities larger than 4,000 inhabitants in 1861) on illiteracy rates in 1911. We consider residualized illiteracy rates after accounting for province fixed effects.

We digitize the first available municipal-level data on illiteracy, collected with the 1911 Census, to assess the long-term impact of the reform on literacy rates, 40 years after its enactment. While we use 1911 data due to availability, this date also allows for a sufficient portion of the population to have been exposed to the reform, while excluding future unexposed cohorts from 1904 onwards.[7] Our findings show that illiteracy rates were 1.6% lower in treated municipalities, representing a 4.8% drop from baseline. This is the first-stage impact of the reform: extending elementary schooling by two additional years significantly reduced analphabetism in treated areas.

[7]On the other hand, the lack of education data prior to 1911 prevents us from employing an instrumental variable specification that would use the reform as an instrument for education since we would be unable to capture the actual impact of the reform on education at the same time as surveillance.

4.2 Empirical Strategy

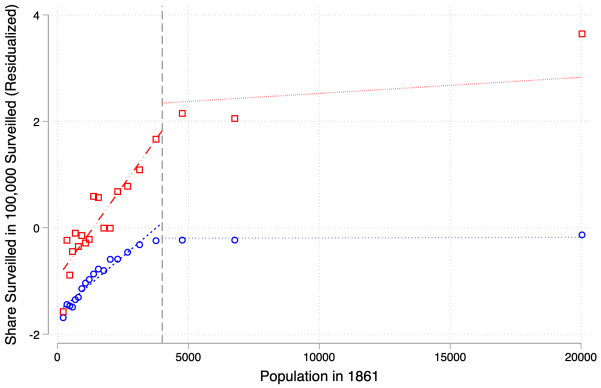
We identify the impact of the education reform on surveillance using a difference-in- discontinuity design (Grembi, Nannicini and Troiano, 2016). We leverage two exogenous sources of variation: discontinuous cross-sectional variation in the municipalities targeted by the reform, and temporal variation in the cohorts subject to the reform. The regression discontinuity component takes advantage of the population cutoff at which the reform intensifies its effect – in municipalities with more than 4,000 inhabitants, where the reform involves an additional two years of schooling. The temporal component leverages variation in cohorts exposed to the education reform due to being seven years old or younger when the reform was enforced, moving them from potentially exiting the education system to staying in school for longer. For municipality i and cohort t, we estimate:

$$\begin{aligned} Surveilled_{it} = \beta\ (Post \times Above4000_{it}) + \lambda(f(Pop1861) \times Post_{it}) \\ + \eta\ (f(Pop1861) \times Post \times Above4000_{it}) + \alpha_i + \theta_t + \kappa_{r \times t} + \epsilon_{it} \end{aligned} \tag{1}$$

where $Surveilled_{it}$ is the population share of individuals surveilled in each municipality-cohort, $Post$ is a dummy taking value 1 for cohorts born after 1854 (seven years old or younger in 1861), $Above4000$ is a dummy equal to 1 for municipalities above 4,000 inhabitants in 1861, and $Pop1861$ is the running variable. All regressions control for municipality, year of birth, and reign r times year of birth fixed effects.[8] The fixed effects account for differences over time, such as the changing methods to collect information on surveilled, differences in municipality characteristics, and for time-varying regional differences. The main coefficient of interest is β, measuring the difference in surveillance due to the education reform at the cutoff. The difference-in-discontinuity component controls for the flexible impact of treatment across population size, allowing us to isolate the effect of a possible jump caused

[8]We consider reign instead of region because our sample starts at a time when Italy was divided in reigns.

Figure 3: Illustration of DinDisc Design



Notes: The figure illustrates our estimation strategy. The number of surveilled per 100,000 inhabitants (residualized by birth year, municipality, and reign-birth year fixed effects) changes at the population cutoff of 4,000 inhabitants for municipality-cohorts born after 1854 and thus exposed to the reform (red), while it does not change for unexposed cohorts (blue).

by the reform at the population discontinuity. We use a first-order polynomial as a baseline but consider other orders for robustness. Standard errors are clustered at the municipality level.

Figure 3 illustrates our difference-in-discontinuity strategy. We separately plot the residualized share of surveilled in municipality-cohorts unexposed to the reform (born before 1854, blue) and exposed to the reform (born after 1854, red) across the population discontinuity of 4,000 inhabitants. Our strategy estimates the impact of the reform by comparing the same cohort of individuals in two municipalities with slightly larger and slightly smaller populations. In line with our expectations, the figure shows that, while no clear discontinuity in surveillance is observed for cohorts unexposed to the reform, there is a positive increase in the share of surveilled individuals among exposed cohorts, providing a preliminary visualization of our findings.

Identifying Assumptions: We provide evidence in favor of the validity of this approach in multiple ways. First, the manipulation test reveals no evidence of sorting in the running variable (Figure C.1). Second, we rely on geographic and climatic covariates to test the credibility of the continuity in potential outcomes assumption. We observe a high level of

balance in covariates across the discontinuity threshold, with none of the differences in means reaching statistical significance (Table C.1). As discussed, there is no other law that changes discontinuously at the 4,000 cutoff, in line with the assumption that the treatment of interest is the only time-varying change at the threshold. We present evidence consistent with the lack of pre-trends in the event-study in the next section.

5 Effect of the Education Reform on Surveillance

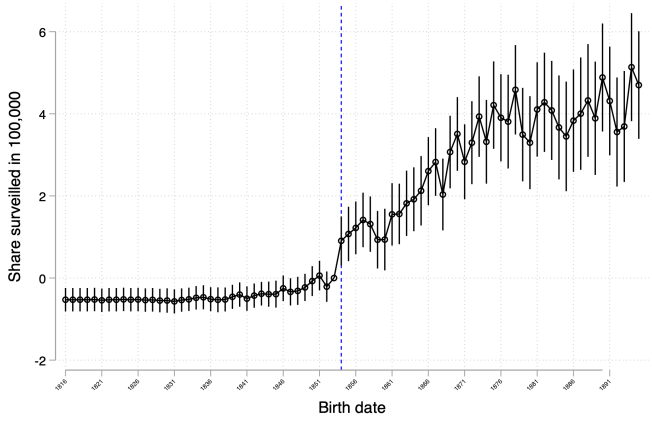
Our main question is whether increased schooling led to a higher probability of a municipality-cohort being targeted by state surveillance. In Table 1, we present our main results. Municipality-cohorts exposed to the education reform experience an additional 3.5 to 3.7 more individual surveilled in 100,000 inhabitants compared to control group municipalities before the reform. This effect is larger when we consider that we start from a baseline surveillance of 0.0824 per 100,000 inhabitants. Assuming 45 cohorts were alive at the same time per municipality[9], the number of surveilled in a municipality would pass from 3.7 to 166 in 100,000 inhabitants per municipality. Effects are similar in magnitude across specifications that do and do not include fixed effects, in line with the possibility that we are isolating the impact of a quasi-random treatment.

We present this analysis as an event study in Figure 4. In the forty cohorts preceding 1854, there was no significant difference in surveillance between municipality-cohorts that will later be exposed to the reform and those that will not. The unusually long pre-period in this setting allows us to provide stronger evidence for the plausibility of the parallel trends assumption than is typically seen in difference-in-differences studies.

Beginning with the 1854 cohort, we observe a stark increase in surveillance, corresponding to an additional 0.4 individuals surveilled per 100,000 in the first year after the reform's implementation. This effect grows over time, reaching an additional 4 individuals surveilled

[9]Based on life expectancy in 1910 in Italy (Caselli, 1990)

Figure 4: Effect of the Education Reform on Surveilled in 100,000 (Event Study)



Notes: The figure plots the effect of the education reform on the number of surveilled per 100,000 inhabitants. We use the specification in Equation 1 for each year of birth. The absorbed category is 1853 ($t-1$ before treatment). Our treatment indicator, *Above*4000 takes value 1 for all cities with more than 4,000 inhabitants in the 1861 Census. The regression includes municipality, date of birth, and reign times date of birth fixed effects. The standard errors are clustered at the municipality level.

per 100,000. The increasing impact of the reform over time is consistent with the Italian state's gradual expansion of capacity to enforce mandatory education, and the broader uptake of educational resources, as schooling became less of an elite privilege. It is likely that the most disadvantaged areas of Italy were slower to benefit from this expansion –illiteracy rates, for example, remained considerably higher in the South even in the 1911 Census. While detailed time-varying data on the expansion of schooling services is absent, preventing us to provide direct evidence on when and where education became more inclusive, we account for fine-grained location fixed effects to account for long-term differences between municipalities, such as poverty or persistently lower provision of public goods.

End of the discontinuity: What happens when the discontinuity in education disappears? As discussed in Section 4.1, compulsory schooling was extended to all municipalities from 1904, eliminating a distinction between places with more and less than 4,000 inhabitants. If the increase in surveillance was produced by the education reform, we should see it disappearing when the distinction between cities is eliminated. In the extended event study in Figure D.1, we see that surveillance decreases exactly for individuals born from 1896 or later

Table 1: Effect of the Education Reform on Surveilled in 100,000 (DDRD)

	(1)	(2)	(3)
	Share Surveilled in 100,000		
Post × Above4000	3.729***	3.605***	3.515***
	(0.359)	(0.350)	(0.318)
Observations	617,040	617,040	617,040
Birth FE	Y	Y	Y
Municipality FE	N	Y	Y
Reign x Birth FE	N	N	Y
Mean DV	0.0824	0.0824	0.0824

Notes: Difference-in-discontinuity results of the effect of the education reform on the number of surveilled per 100,000 inhabitants in a municipality-cohort (Equation 1). Our coefficient of interest is *Post* × *Above*4000. The variable *Post* takes value 1 for cohorts born from 1854 onwards (seven years old or younger when the reform was enacted). The variable *Above*4000 takes value 1 for municipalities with more than 4,000 inhabitants in the 1861 census. We report the mean of the dependent variable for cities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

in municipalities above 4,000 inhabitants, now exposed to the same education requirements as other municipalities. As for the main result, the difference between previously treated and control municipalities does not disappear immediately: places that had built school capital over time maintain an advantage in effectively providing education compared to newly treated locations, resulting in a slowly decreasing impact on surveillance.

Robustness: These findings are consistent across a variety of robustness tests. We begin by considering alternative transformations of the dependent variable in Table D.2, replacing the share of surveilled in 100,000 inhabitants with a dummy equal to one for whether any individuals born in the cohort within the municipality was surveilled (column 1). At the threshold, there is a 28% higher likelihood of surveillance per municipality-cohort after the

reform. In columns 2 and 3 we consider the absolute number and the log absolute number of individuals surveilled, finding a significant effect only with a logarithmic transformation of the dependent variable, consistently with this being highly skewed towards zero. Table D.3 considers higher order polynomials for the interaction of population, treatment, and post, controlling flexibly for the impact of population on either side of the discontinuity. Table D.4 tests robustness to additional specifications, including controlling for the date of surveillance (column 1), using a Poisson specification (column 2), and accounting for the reforms' late implementation in areas that were annexed to the Kingdom of Italy after 1861 (column 3).

Double DDRD and DiD: In Table D.5, we consider the effect of the reform when controlling for discontinuities across both population and cohort of birth by flexibly interacting not only population but also date of birth with each term in Equation 1. This fully saturated model is effectively the equivalent of a double difference-in-discontinuities. Our results are almost identical in magnitude when considering this additional set of control variables. Instead, in Table D.6 we replicate our findings with a simpler and less restrictive specification: a difference-in-differences framework, where we do not flexibly control for distance to the population discontinuity. Again, findings are similar in significance and magnitude with this specification.

Placebos: We present two sets of placebo analyses to rule out the possibility that other factors cause the jump we observe at the education reform discontinuity. First, we test the impact of other population threshold, above or below the actual discontinuity. We find that none of those is significant in explaining surveillance in Tables D.9 and D.10. This placebo also responds to a possible concern that governments might pick larger municipalities to pay more attention to by looking at locations above a certain population number. Since all the placebos we consider are round numbers and we see no effect at those, we can discard this possibility.

Second, we repeat the same exercise for cohort of birth and ask whether being born in

another cohort unexposed to the reform is likely to trigger the same effect by replacing our *Post* indicator with a set of placebo cohorts. Findings presented in Table D.5 rule out the importance of non-treated cohorts in explaining the effect we find on surveillance.

6 The Role of Social Class

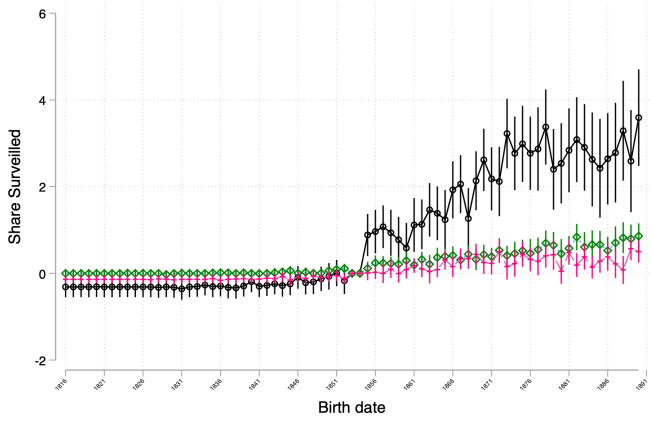
The education reform led to increased surveillance among treated cohorts. To assess whether this rise in surveillance is connected to state concerns over the education of marginalized populations, we classify the professions of surveilled individuals into distinct social classes, as detailed in the Data section and illustrated in Figure B.3. We then use the same specification as in Equation 1 on three distinct dependent variables: surveilled from the working, middle, and upper class. The results, presented in both event-study plots (Figure 5) and table format (Table 2), indicate that the increased surveillance following the reform is primarily driven by the working class. The increase in surveillance for this group is large: there are 2.5 more individuals surveilled in 100,000 –almost a 45-fold increase from the mean of the dependent variable at baseline. We observe a positive but smaller effect for the middle and upper classes, corresponding to a 0.49 and 0.37 increase respectively. The differential increase in surveillance for working class individuals is also visible in the event study plot.

Table 2: Effect of the Education Reform on Surveilled in 100,000 by class (DDRD)

	(1)	(2)	(3)
	Share Surveilled in 100,000		
	Working class	Middle class	Upper class
Post × Above4000	2.488***	0.475***	0.365***
	(0.241)	(0.0504)	(0.0439)
Observations	617,040	617,040	617,040
Birth FE	Y	Y	Y
Municipality FE	Y	Y	Y
Reign x Birth FE	Y	Y	Y
Mean DV	0.0518	0.0115	0.0122
Δ over mean	48.03	41.31	29.91

Notes: DDRD results (Equation 1) of the effect of the education reform on the number of surveilled per 100,000 inhabitants in a municipality-cohort by class.

Figure 5: Effect of the Education Reform on Surveillance by Class (Event study)



Notes: DDRD results (Equation 1) of the effect of the education reform on the number of surveilled per 100,000 inhabitants in a municipality-cohort. The dependent variable is surveilled from the working (black), middle (green), and upper class (pink). The absorbed category is 1853.

A potential concern with this analysis is that the number of individuals from the working class might itself be influenced by the education reform. However, this is not an issue for two reasons. First, our results remain consistent and, in fact, become larger when we use a binary indicator for the presence of any surveilled individuals from that social class (Table E.1). For the working class, the increase is 21.6 percentage points (a 16-fold increase from the control mean), while for middle and upper classes the increase is 3 percentage points (an 8 and 7-fold increase). Second, if the reform expanded access to literacy and improved the labor prospects of low-income treated cohorts, it could have reduced the overall proportion of individuals classified as working-class in the post-reform period. This would imply that our findings represent a lower bound of the true effect: we cannot directly observe the fraction of the population reclassified as non-working-class due to improved opportunities. Therefore, given that the share of the working-class population likely declined in treated municipalities, the observed effects of surveillance in this analysis should be interpreted as a conservative estimate of the actual impact.

7 Individual Activation does not Explain Surveillance

Our findings on surveillance could be the result of two forces: states being concerned with the political activity of individuals in low socioeconomic classes, and greater political mobilization triggered by increased education. We employ five indicators of political activism to consider this possibility, ranging from the least to the most costly forms of political action.

Voter mobilization: First, we consider the impact of the reform on voter mobilization using national elections starting from the first election in unified Italy, held in 1861. We adopt the difference-in-discontinuities approach in Equation 1 to measure the impact of the education reform on voter turnout, but unlike in the main analysis, here we consider a database at the municipality-election year level. In this case, our first treated year is election year 1880, when individuals first treated by the reform (born in 1854) turn to voting age (25 years old). Before the first affected cohort starts voting, municipalities above the 4,000-population threshold voted at about the same rate as those below (as shown in a RDD in Figure F.1). This continuity in casted votes persists after the treated cohorts gain the right to vote (Table F.1 and Figure 6, panel (a)). The same finding is confirmed if we plot the coefficient of the regression discontinuity design estimated cross-sectionally on each election year. Also in this specification, we observe flat pre-trends in votes cast and no differential change afterwards (Figure F.2).

Notable people involved in politics: Our second measure of political engagement is the number of individuals involved in politics. We scrape information from Wikidata on all notable historical figures born in Italy from 1816 to 1896 – the same time period as the surveillance data – and we code those who held any type of political roles, including political activists, diplomats, senators, suffragists, but also bishops, who effectively held political roles. This data allows us to replicate the same municipality-cohort specification as in our main analysis on surveillance. As evidenced in Table F.2 and Figure 6, panel (b), there is a

null or negative impact of the reform on individuals in any role, including both political and non-political positions.

Protests and strikes: Third, we consider a medium-level form of activism, protests and strikes. We scrape news items on protests and strikes from a newspaper with archives dating back to the Kingdom of Italy, La Stampa. The location the news item refers to is either reported by the newspaper or, when it is not, we search municipality names in the article and geocode the news as happening in cited municipalities, excluding municipalities with names having double meaning. In this case, we do not have a pre-period before the reform. As a result of not being able to exploit the temporal difference-in-difference component, we rely only on the cross-sectional regression discontinuity design at 4,000 inhabitants. Consistently with the previous results, our findings indicate no significant increase in politically motivated activities in treated areas (Figure 6, panel (d) and Table F.4).

Partisans and Bandits: Finally, we consider two extremely intensive forms of political activism: participating in the resistance movement and engaging in banditry. Partisans were guerrilla fighters resisting Fascist and Nazi forces through sabotage and combat. Bandits, often seen as rebels rather than criminals, carried out raids and kidnappings, particularly in southern and central Italy, where social inequalities were severe (Ciconte, 2011). Both were armed forms of action involving a high risk of death. The data on partisans includes date and location of birth, allowing us to run the same DDRD specification as in the main analyses. When we use the share of partisans (as for our main specification on surveillance), we find a null effect of the education reform on the likelihood of engaging in the resistance movement (Table F.5, column 1). Effects are mixed with alternative dependent variables – they become negative considering the total number of partisans, and positive considering a dummy. Figure 6, panel (c), instead, shows a clear null effect on the number of bandits.[10]

[10]The cross-sectional nature of the data prevents us from running a DDRD.

Does political engagement drop because of surveillance? While we find no evidence of heightened political engagement, it could still be the case that these null and negative results are directly caused by higher surveillance. On one hand, it was shown that totalitarian surveillance in East Germany and authoritarian control in China depressed political engagement (Hager and Krakowski, 2022; Xu, 2021). On the other hand, in a setting where the state was not authoritarian and repression was not large-scale, and where only the most engaged individuals were likely aware of being under police scrutiny, participation might have suffered less from increased state surveillance. We consider this possibility in two ways.

First, we test the impact of the education reform on political engagement for cohorts that were not yet surveilled. We perform this analysis on notable political figures, as this is the only metric of political engagement for which we have cohort-level data. If surveillance had a systematic impact on political participation, we would observe *more* political figures in cohorts subject to the education reform but where surveillance was absent. Instead, we observe again a null effect on the number of political figures in these cohorts, and an effect that is not differential from that on other cohorts exposed to surveillance (Table F.2).

Second, we run a heterogeneity analyses for each indicator of political engagement by whether the level of surveillance was high or low (Table F.3). Across analyses, we find no differential impact of high surveillance on any metric of political activism. While both these analyses are not causal – in the first case, there might be endogenous reasons why certain cohorts are not surveilled, and the second analysis is a simple heterogeneity – all results point to the same conclusion of no differential impact of the education reform on political engagement depending on surveillance.

These findings align with the possibility that, outside of contexts where the state is highly repressive and individuals expect to be subject of tight surveillance, as in totalitarian regimes, surveillance may not have a large impact engagement at least not to a degree detectable in our analyses.

Why did the education reform impact surveillance but not political engagement?

State surveillance is driven by perceived threats rather than by the number of individuals mobilized. Officials may have assumed that educated members of the working class were more capable of political rebellion, prompting preemptive monitoring. However, the actual number of individuals politically activated by education within the working class may have remained too low to detect it statistically. Economic and institutional barriers likely constrained their ability to translate alphabetism into action at scale. Yet, even a small number of individuals showing signs of mobilization or merely possessing the potential to do so could have been sufficient to trigger increased state surveillance. The lack of evidence of increased political engagement following the reform, supports the possibility that heightened surveillance was not primarily driven by increased political mobilization but rather by state concerns over newly educated, potentially influential working-class citizens.

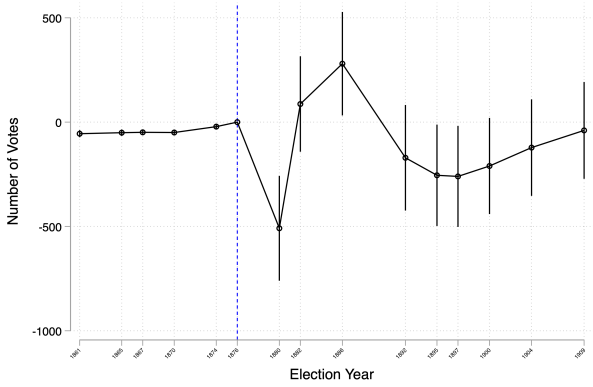
8 State Activation as a Determinant of Surveillance

Up to this point, we have shown that the education reform led to increased surveillance, particularly among individuals in working-class occupations. We have also shown, considering a wide array of indicators, that this effect is not driven by a general rise in political mobilization. We now provide evidence for the other candidate explanation: that the increase in surveillance reflects state concerns about marginalized individuals acquiring political skills through education.

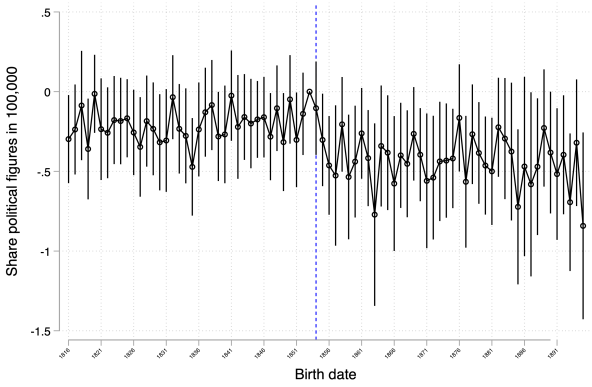
Two observable implications follow from this hypothesis. First, if the state perceives this group as a threat, they should be subjected to more intensive surveillance. The next subsection evaluates three indicators of surveillance intensity: the severity of measures, the duration of surveillance, and the volume of information collected. Second, surveillance should be concentrated on individuals advocating for marginalized groups. We examine this possibility by considering the ideology of those under surveillance. Lastly, we consider heterogeneity analyses for factors unlikely to change the use of this strategic behavior –

Figure 6: Effect of the education reform on political engagement

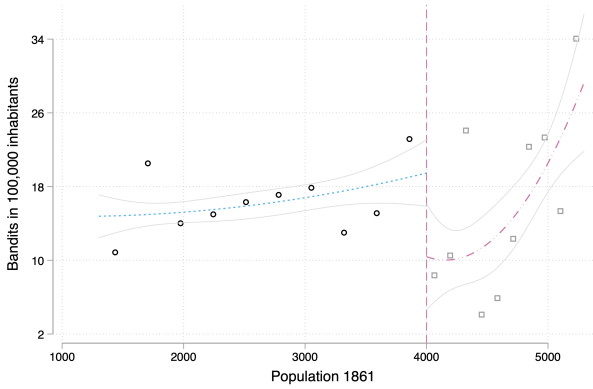
(a) Effect on number of votes (DDRD)



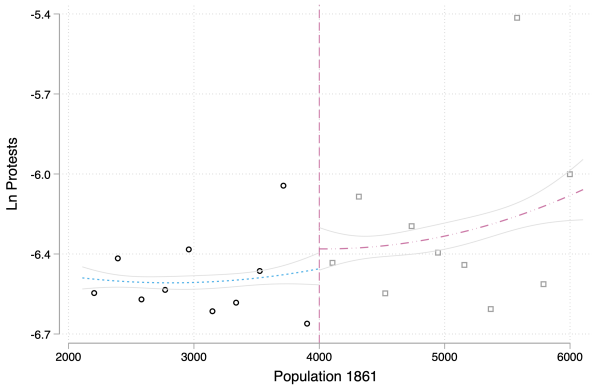
(b) Effect on political figures (DDRD)



(c) Effect on bandits (RDD)



(d) Effect on protests (RDD)



Notes: (a) Eventstudy of the effect of the reform on the number of votes per election year. We use the specification in Equation 1 for each election year. Individuals who were 7 years or younger at the time of the reform voted for the first time in 1880. The absorbed category is 1876. The regression includes municipality, election year and reign times year fixed effects. Standard errors are clustered at the municipal level. (b) Eventstudy of the effect of the reform on the share of notable political figures in 100,000 inhabitants in a municipality-cohort. We use the specification in Equation 1 for each year of birth. The regression includes municipality, date of birth and reign time date of birth fixed effects. Standard errors are clustered at the municipal level. (c) Effect of being in populations above the 4,000 cutoff on the share of bandits in 100,000 inhabitants. Standard errors are clustered at the municipal level. (d) Effect of being in populations above the 4,000 cutoff on the share of protests in 100,000. Standard errors are clustered at the municipal level.

regime type and state capacity.

8.1 Intensity of Surveillance

Severity: First, we use code whether individuals were under severe forms of surveillance, including political confinement, a costly tool used to monitor dissidents' movements and in some cases to forcibly remove them from their enclaves. Table E.2, column 1, shows that municipality-cohorts affected by the education reform experienced higher rates of severe surveillance measures. The effect is concentrated among the working class (column 2) and is less pronounced for other groups (columns 3 and 4).

Duration: The second measure considers the length of time an individual remains under surveillance. On average, a given municipality-cohort is surveilled for 14 days per year. However, in treated cohorts, the duration of surveillance increases by more than a factor of 100 (Table E.3, column 1). As with the severity of surveillance, this effect is most pronounced among the working class, where surveillance duration increases by a factor of 185 (column 2), exceeding the increase observed for middle and upper-class individuals (columns 3 and 4).

Quantity of Information: The third measure examines whether authorities collected extensive information on the surveilled, as indicated by the presence of a biographical file ("scheda biografica"). These files were maintained for all significant political figures or for individuals considered particularly dangerous. Descriptive statistics show that such files were more frequently recorded for working-class individuals (2.3%) compared to middle-class (1.8%) and upper-class (0.9%) individuals. Table E.4, column 1, shows that treated cohorts were more likely to have a biographical file, with the share of surveilled individuals with biographical documentation increasing by 26% over the baseline mean. This effect is, again, strongest among the working class, with an increase of 29% (column 2), larger than for any other class (column 3 and 4).

Across all three metrics, the data indicate that the state implemented more intensive

surveillance measures against individuals affected by the education reform, particularly those in working-class occupations. These findings are consistent with the argument that state fears drive surveillance practices.

8.2 Ideology of Surveilled

Having established how the state varies the intensity of surveillance to target certain groups, the analysis now turns to the ideological orientation of those targeted. If surveillance reflects state concerns about educated but economically marginalized groups, individuals advocating for the rights of economically disadvantaged groups should be disproportionately monitored.

Table D.8 presents surveillance rates by ideological affiliation. While all political groups experienced increased monitoring following the education reform, socialists and communists were surveilled at the highest rates. This pattern aligns with their ideological focus on addressing the demands and grievances of marginalized classes. Figure D.3 further disaggregates these results by social class, showing that among the four ideological categories examined, the surveillance gap between the working class and other groups is largest for socialists and communists. These findings are consistent with the earlier results on surveillance intensity, indicating that the state disproportionately targets individuals advocating for proletarian rights, particularly those from the working class.

8.3 Additional Heterogeneities

Democracy and Dictatorship: A substantively important question is whether the strategic use of state surveillance we uncover in these analyses varies across regime types. To answer this question, we leverage our unique setting, which allows us to observe state behavior across both democratic (1869-1921) and fascist governments (1922-1945). Table D.7 presents results separately for these two periods. The findings indicate similar patterns across both regimes. Under democracy (columns 1-4), the likelihood of surveillance increases by a factor of 1.8, while under fascism (columns 5-8), it increases by a factor of 1.6. In both cases, the effect

is largest among the working class, with increases of 1.37 (democracy) and 1.125 (fascism). The consistent patterns across regimes suggest that educated yet economically marginalized individuals remain a primary focus of state surveillance, regardless of the governing ideology.

State Capacity: Another potential factor influencing surveillance intensity is state capacity. To assess this, we digitize new data on postal offices in the Kingdom of Italy, sourced from an 1861 official postal bulletin, as a proxy for state presence. Table D.14 shows that surveillance is more prevalent in areas with postal offices, but the effect remains significant even in their absence. Second, Table D.13 examines geographical variation in treatment effects, revealing that while the impact of education reform on surveillance is stronger in Central Italy, similar patterns are observed across all regions. These results indicate that, although greater state capacity facilitates surveillance, it is not a prerequisite for the patterns documented here.

Pre-Reform Surveillance Levels: Finally, the analysis considers whether the increase in surveillance is driven by newly monitored areas or by heightened scrutiny in already highly surveilled regions. Table D.12 shows that both channels contribute to the observed effects. Treatment effects are larger in areas with higher pre-reform surveillance, possibly due to spillover from surveilling more individuals, or network-driven mobilization. However, the effect remains significant even when accounting for pre-reform surveillance levels.

9 Alternative Explanations

Migration of educated individuals: Our specification uses individuals' birthplace to determine their exposure to the education reform during elementary school. However, it is possible that educated individuals were more likely to migrate to larger cities for better opportunities, where they might encounter heightened political activism. In this scenario, increased surveillance could stem from education-driven migration, as individuals with higher education gravitate toward politically active urban centers. Fortunately, our data include

both the birth and residence locations of surveilled individuals, allowing us to test this possibility by limiting the analysis to those who remained in their birthplace. Results in Table G.2 confirm our primary findings even among non-migrants: treated municipality-cohorts experience an increase in the number of surveilled individuals to about 40 times the baseline –a result consistent with those shown in Table 1. Therefore, while education may indeed have influenced migration and increased exposure to activism, this does not appear to be the main driver of our results.

Teachers as state agents: One possibility is that teachers, as part of the bureaucratic system, may have acted as informers for the state, monitoring former students and report on their political activity. While we cannot directly test this due to the absence of data on informants, there are two reasons to doubt this scenario. First, historical records indicate that informants were generally individuals in dire personal situations and working as informants to attenuate personal problems, such as prisoners seeking reduced sentences by reporting on their jail mates (Ciconte, 2011). Instead, teachers held stable, financially stable, and high status positions, far from the situations of captivity that motivated informants. Second, if state bureaucrats were used as informants, we would expect that also other reforms increasing the size of the bureaucracy would lead to a similar surge in surveillance. To test this, we examine the impact of the Crispi Law of 1891, which required municipalities with over 5,000 residents to expand their welfare boards, thereby substantially increasing the administrative workforce. Results in Table G.1 reveal that this expansion had no measurable effect on the level of surveillance. This indirect test reinforces the historical accounts, suggesting that an increase in the number of bureaucrats alone did not necessarily lead to heightened state surveillance.

Teachers as instigators: Another possibility is that the state perceives teachers as dangerous instigators of radical ideologies. If this was the case, we would observe that the increase in surveillance is driven by a surge in surveillance of teachers, which are a larger

group in treated areas. First, teachers made up a negligible fraction of those surveilled pre-reform (0.00005 per 100,000 inhabitants), making it improbable that their surveillance alone drives the treatment effect. Second, our analysis shows that the rise in teacher surveillance (column 2) is smaller than for other professions with similar baseline surveillance (columns 3–5, Table G.3), despite the post-reform increase in teachers' share of the population. Moreover, the absolute number of teachers surveilled does not increase (column 1).

Substitution with the Church: One effect of the education reform is that schools began to take on the church's role in education, as prior to the reform, churches were largely responsible for occupying and educating children. However, since our analysis centers on rebellion against the state, and the reform positioned the state as the provider of education, it is implausible that state-led schooling would inspire greater rebellion against itself. In fact, research by Paglayan (2024) shows that states typically design education reforms to cultivate loyalty and compliance among citizens. Therefore, if a substitution from church to state oversight occurred, as seems likely, we would expect the opposite effect: with increased exposure to state-run education, surveillance should have decreased, not increased.

Education of other actors: Could the increase in surveillance be linked to increased education among state actors, such as the police, rather than the surveilled individuals themselves? This seems unlikely for two reasons. First, the secret services were a highly educated section of the police which would have been unlikely to change its education status after the reform. Second, for this alternative story to explain the sudden increase in surveillance in the cohort post-reform, two critical conditions must be met: (i) the increased surveillance would need to stem precisely from the new cohort of educated police, and (ii) these police would need to disproportionately target individuals of their exact same age cohort, as the increase in surveillance is driven by the age of those surveilled rather than the year of surveillance. This specific pattern of age-driven targeting lacks institutional rationale and might be too demanding to be plausible.

10 Conclusions

Our study sheds new light on the strategic considerations states make when deciding whom to subject to political surveillance. Using a unique dataset on Italian political dissidents covering over a century, we examine the impact of an education reform that targeted lower socioeconomic groups, increasing their access to schooling. Our findings reveal that the reform led to an increase in state surveillance, particularly for individuals from working-class backgrounds. This effect is valid across democratic and authoritarian regimes, and it is not driven by migration patterns or bureaucratic expansion. We also find limited evidence that an increase in participation can explain these effects, indicating that increased surveillance spurs from state heightened concern rather than from an effectively higher threat. Our findings support a view in which states perceive educated but economically marginalized individuals as particularly threatening to regime stability, regardless of whether these individuals engage in overt political activism.

To what extent do historical findings on surveillance remain relevant in the age of digital monitoring? Specifically, do states still face resource constraints that force them to choose whom to surveil, or have digital tools eliminated the need for selectivity, enabling widespread and indiscriminate targeting? On one hand, social media and facial recognition technologies have made it easier for states to gather information on individuals. On the other hand, technological advancements also provide activists with new ways to evade scrutiny, such as encrypted communication and an abundance of alternative channels for correspondence. Moreover, while digital tools have expanded surveillance capabilities, collecting, storing, and analyzing vast amounts of data remains costly, particularly when it requires transforming unstructured information into actionable intelligence. In authoritarian contexts, where dissidents are acutely aware of surveillance risks, large datasets may yield little useful intelligence, as activists take steps to avoid leaving digital traces of their actions. As a result, states may still need to rely on traditional intelligence-gathering methods. Considering these factors, we argue that it remains useful to think of states as resource-constrained actors

who must make strategic choices about whom to monitor. This perspective reinforces the relevance of our findings for understanding political surveillance in the digital age.

This research offers a new perspective on the enduring challenges faced by states in balancing security with democratic inclusivity. As states around the world continue to adopt increasingly sophisticated surveillance technologies, our findings emphasize the need to ensure that these tools do not disproportionately target or disenfranchise already vulnerable populations.

References

Acemoglu, Daron and James A Robinson. 2001. "A theory of political transitions." *American Economic Review* 91(4):938–963.

Almond, Gabriel Abraham, Sidney Verba et al. 1963. "The civic culture: Political attitudes and democracy in five nations.".

Andirin, Veli, Yusuf Neggers, Mehdi Shadmehr and Jesse M Shapiro. 2022. Real-time Surveillance of Repression: Theory and Implementation. Technical report National Bureau of Economic Research.

Beraja, Martin, Andrew Kao, David Y Yang and Noam Yuchtman. 2023. "AI-tocracy." *The Quarterly Journal of Economics* 138(3):1349–1402.

Blaydes, Lisa. 2018. *State of Repression: Iraq under Saddam Hussein*. Princeton University Press.

Bourguignon, François and Thierry Verdier. 2000. "Oligarchy, democracy, inequality and growth." *Journal of development Economics* 62(2):285–313.

Bueno de Mesquita, Bruce and Alastair Smith. 2017. "Political succession: A model of coups, revolution, purges, and everyday politics." *Journal of Conflict Resolution* 61(4):707–743.

Calonico, Sebastian, Matias D Cattaneo and Rocio Titiunik. 2015. "Optimal data-driven regression discontinuity plots." *Journal of the American Statistical Association* 110(512):1753–1769.

Carolini, Simonetta and Fabio Ecca. 2015. *Antifascisti alla sbarra*. Palombi.
URL: *https://anppia.it/gli-antifascisti/*

Caselli, G. 1990. "Mortalità e sopravvivenza in Italia dall'Unità agli anni'30." *Popolazione, Società e ambiente. Temi di Demografia storica italiana (secc XVII-XIX)* pp. 275–310.

Ciancarini, Enrico. 2019. "Appunti di vite sovversive." *Società Storica Civitavecchiese* .

Ciconte, Enzo. 2011. *Banditi e briganti: una storia senza fine, dal Cinquecento all'Ottocento*. Rubettino.

Davenport, Christian. 2005. "Understanding covert repressive action: The case of the US government against the Republic of New Africa." *Journal of Conflict Resolution* 49(1):120–140.

Di Fiore, Laura. 2018. *Gli Invisibili: Polizia politica e agenti segreti nell'Ottocento borbonico*. Vol. 21 FedOA-Federico II University Press.

Dilemmi, Andrea. 2010. Polizia e sorveglianza del dissenso politico (Verona, 1894-1963) PhD thesis Università di Verona. Doctoral thesis in Scienze storiche e antropologiche.

Esberg, Jane. 2021. "Anticipating dissent: The repression of politicians in Pinochet's Chile." *The Journal of Politics* 83(2):689–705.

Fanon, Frantz. 1963. "The Wretched of the Earth." *Grove Weidenfeld* .

Garfias, Francisco and Beatriz Magaloni. 2018. "Taking to the streets: Protests and regime survival." *no journal* .

Gramsci, Antonio. 1975. *Quaderni del carcere*. Torino: Einaudi. Critical edition edited by Valentino Gerratana.

Grembi, Veronica, Tommaso Nannicini and Ugo Troiano. 2016. "Do fiscal rules matter?" *American Economic Journal: Applied Economics* pp. 1–30.

Gurr, Ted Robert. 1970. *Why Men Rebel*. Princeton, NJ: Princeton University Press.

Hager, Anselm and Krzysztof Krakowski. 2022. "Does state repression spark protests? evidence from secret police surveillance in communist poland." *American Political Science Review* 116(2):564–579.

Hobsbawm, Eric J. 1971. *Primitive rebels: Studies in archaic forms of social movement in the 19th and 20th centuries*. Manchester University Press.

Klor, Esteban F, Sebastián M Saiegh and Shanker Satyanath. 2016. "Crony capitalism and the targeting of violence: labor repression during Argentina's last dictatorship.".

Kuran, Timur. 1989. "Sparks and prairie fires: A theory of unanticipated political revolution." *Public choice* 61(1):41–74.

Labini, Paolo Sylos. 1972. "Sviluppo economico e classi sociali in Italia." *Astrolabio, n.3* pp. 18–31.

Lecce, Giampaolo, Laura Ogliari and Tommaso Orlando. 2022. "State formation, social unrest and cultural distance: Brigandage in post-unification Italy." *Journal of Economic Growth* 27(3):453–483.

Leto, Guido. 1951. *OVRA: Fascismo-Antifascismo*. Bologna: Cappelli.

Lichter, Andreas, Max Löffler and Sebastian Siegloch. 2021. "The long-term costs of government surveillance: Insights from Stasi spying in East Germany." *Journal of the European Economic Association* 19(2):741–789.

Marx, Karl and Friedrich Engels. 1848. *Manifesto of the Communist Party.* Workers' Educational Association. Reprinted in various editions.
URL: *https://www.marxists.org/archive/marx/works/1848/communist-manifesto/*

Nugent, Elizabeth R. 2020. "The psychology of repression and polarization." *World Politics* 72(2):291–334.

Paglayan, Agustina. 2024. *Raised to Obey: The Rise and Spread of Mass Education.* Vol. 134 Princeton University Press.

Pan, Jennifer and Alexandra A Siegel. 2020. "How Saudi crackdowns fail to silence online dissent." *American Political Science Review* 114(1):109–125.

Pierskalla, Jan Henryk. 2010. "Protest, deterrence, and escalation: The strategic calculus of government repression." *Journal of conflict Resolution* 54(1):117–145.

Ritter, Emily Hencken and Courtenay R Conrad. 2016. "Preventing and responding to dissent: The observational challenges of explaining strategic repression." *American Political Science Review* 110(1):85–99.

Rozenas, Arturas and Yuri M Zhukov. 2019. "Mass repression and political loyalty: Evidence from Stalin's 'terror by hunger'." *American Political Science Review* 113(2):569–583.

Scott, James C. 1985. *Weapons of the Weak: Everyday Forms of Peasant Resistance.* Yale University Press.

Shadmehr, Mehdi. 2014. "Mobilization, repression, and revolution: Grievances and opportunities in contentious politics." *The Journal of Politics* 76(3):621–635.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China.* Cambridge: Cambridge University Press.

Sotgiu, Girolamo. 1984. *Storia della Sardegna sabauda 1720-1847.* Roma-Bari: Laterza.

Sullivan, Christopher M. 2016. "Political repression and the destruction of dissident organizations: Evidence from the archives of the Guatemalan national police." *World Politics* 68(4):645–676.

Thachil, Tariq. 2020. "Does police repression spur everyday cooperation? Evidence from urban India." *The Journal of Politics* 82(4):1474–1489.

Tosatti, Giovanna. 1990. "Le origini del Casellario politico centrale." *Istituto per la scienza dell'amministrazione pubblica, Archivio Nuova serie 6 "Le riforme crispine", vol. I "Amministrazione statale"* .

Truex, Rory. 2019. "Focal points, dissident calendars, and preemptive repression." *Journal of Conflict Resolution* 63(4):1032–1052.

Verba, Sidney, Kay Lehman Schlozman, Henry Brady and Norman H Nie. 1993. "Citizen activity: Who participates? What do they say?" *American Political Science Review* 87(2):303–318.

Xu, Xu. 2021. "To repress or to co-opt? Authoritarian control in the age of digital surveillance." *American Journal of Political Science* 65(2):309–325.

Young, Lauren E. 2019. "The psychology of state repression: Fear and dissent decisions in Zimbabwe." *American Political Science Review* 113(1):140–155.

SUPPLEMENTARY INFORMATION

— For Online Publication —

A Additional Details on Data

Voting behavior: Data on the voting behavior of the Italian population was obtained from the Electoral Historical Atlas Archive, which includes the results of all general elections held in Italy from 1861 to 2008. For the period between 1861 and 1913, which is the focus of our analysis, the archive provides records of the total number of votes for each candidate at the municipal level. However, there is no available information on the total number of eligible voters. In our analysis, we constructed a panel dataset at the municipality level, where the key variables are the number of votes cast per 100,000 inhabitants and the absolute number of votes cast.

Wikidata: The information on notable individuals known for their roles in politics was sourced from Wikidata. We scraped records for individuals born in Italy between 1816 and 1930. For each individual, we collected data on their birth date, municipality of birth, occupation, and other significant roles held, when available. Based on the last two fields, we identified notable individuals involved in politics. We classified as political those occupations that identify individuals as activists, partisans, bureaucrats, or religious figures.

Protests: Information on location and timing of protests was sourced from articles published by the Italian newspaper *La Stampa*, the only including articles dating as far back as 1867. We scraped all articles published between 1867 and 1940 using the keywords *protesta* (protest) and *sciopero* (strike). For each article, we extracted and stored key information such as the title, abstract, publication date, and references to specific locations. We assigned each protest to the location mentioned in the article and then used this data to construct a panel dataset. To ensure accuracy, we excluded cities whose names are also common nouns,

such as Porte ("doors"), or Ora ("hour"). For a subset of these common words, we manually reviewed the articles and confirmed that the mentions referred to the common noun, rather than the municipality. Of all the municipalities in our sample, only 0.5% had names that are common words, and of these, only 10 had populations greater than 4,000. These cities were excluded from the protests analysis to avoid misclassification.

A.1 Education Reforms

Our main analysis considers the first education reform of Unified Italy, the Casati Law enacted in 1861. In 1877, another education reform was passed, the Coppino Law. This law introduced new curricula for civic education, imposed penalties for failure to comply with mandatory education, and further extended compulsory schooling by an additional year in municipalities below 4,000 inhabitants. This discontinuity entails that our treatment becomes one (rather than two) additional year of education for individuals that were seven years old or younger in 1877 (e.g. born in 1870).

The discontinuity in education disappears, instead, in 1890 with the Orlando Law, which extended mandatory education until age 12 to all municipalities, independently on their population.

A.2 A Non-Education (Placebo) Reform:

The Crispi Law mandated that local governments assume responsibility for administering public welfare institutions such as orphanages, shelters, and other charitable organizations. The law established local welfare boards which were tasked with overseeing welfare programs to ensure that aid reached vulnerable groups. These boards were subject to rules: in municipalities with populations exceeding 5,000, the boards were required to have eight members, while smaller municipalities were only required to have three. This provision aimed to ensure that larger populations received more robust administrative support.

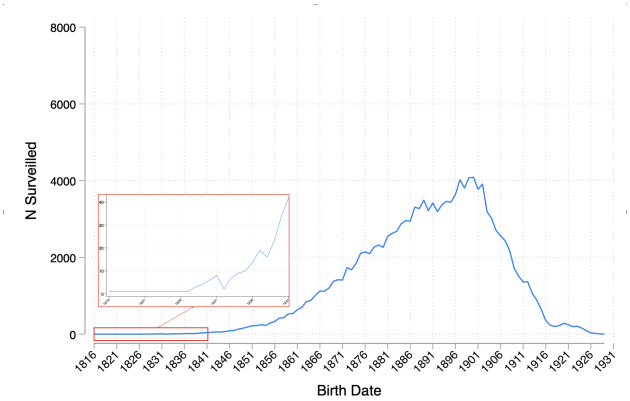
B Descriptives

Figure B.1: Number of surveilled by municipality of birth



Notes: The map shows the total number of individuals surveilled by municipality of birth.

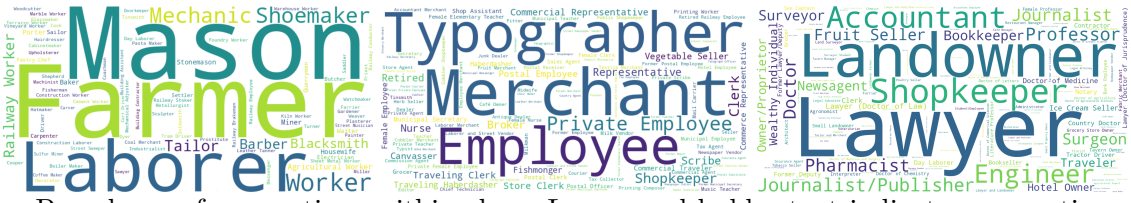
Figure B.2: Number of surveilled by date of birth



Notes: The graph shows the number of individuals surveilled by year of birth. The red rectangle zooms on the earlier decades, where the number of surveilled is significantly lower – but non-zero.

Figure B.3: Jobs distribution within classes

(a) Working Class (b) Middle Class (c) Upper Class



Notes: Prevalence of occupations within class. Larger and bolder text indicates occupations with a higher number of workers.

Figure B.4: Number of surveilled by class and political ideology

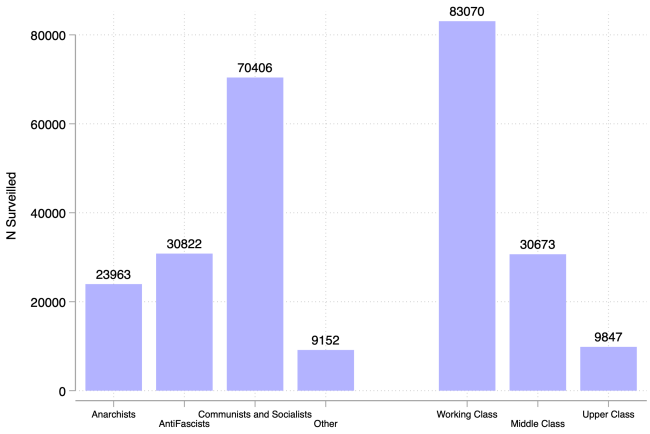
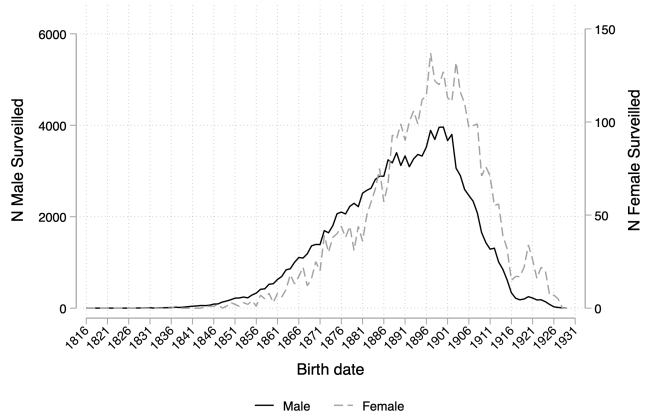
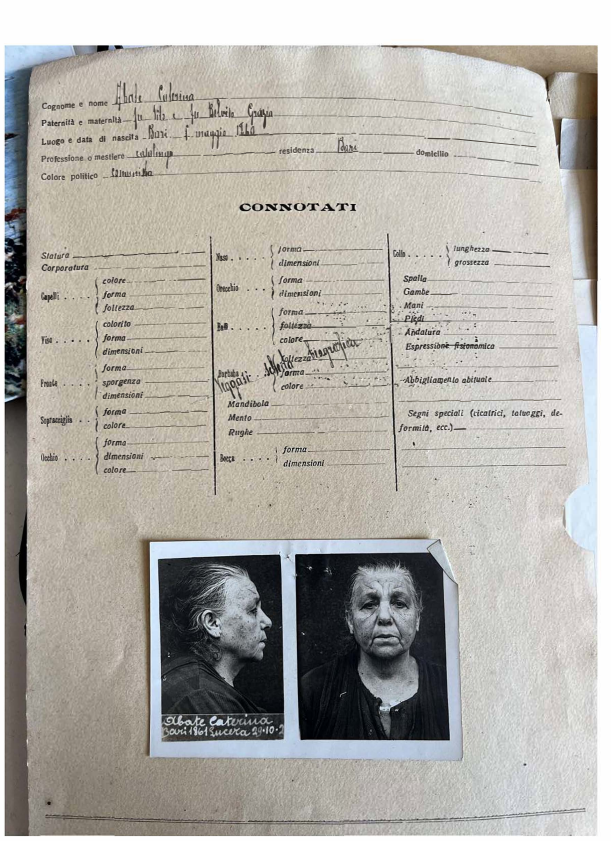


Figure B.5: Number of surveilled by gender



Notes: The figure displays the number of individuals surveilled by gender. In solid black, the number of male surveilled by date of birth (left axis). In dashed gray, the number of female surveilled by date of birth (right axis).

Figure B.6: Example of a biographical file of a surveilled



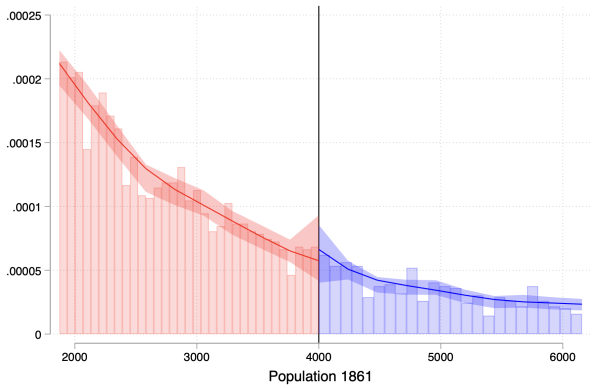
C Identifying Assumptions

Table C.1: Balance Table of the RD sample in fixed and post-treatment outcomes

	Mean T	Mean C	Diff	Std Diff
Minimum Elevation	182.184	159.508	-12.486	-0.095
	(171.413)	(167.527)	(64.936)	
Maximum Elevation	828.189	802.665	-162.232	-0.028
	(641.521)	(641.042)	(245.545)	
Solar radiation (kJ/mˆ2 day)	10,727.552	10,878.235	-448.329	0.083
	(1,248.738)	(1,327.522)	(490.641)	
Minimum Temperature (°C)	9.942	10.273	-0.477	0.092
	(2.588)	(2.521)	(0.981)	
Maximum Temperature (°C)	17.947	18.381	-0.203	0.125
	(2.475)	(2.436)	(0.941)	
Water Vapor Pressure (kPa)	1.248	1.276	-0.044	0.111
	(0.180)	(0.177)	(0.068)	
Bioclimatic Variables	21.184	21.473	0.026	0.099
	(2.131)	(2.012)	(0.797)	
Precipitation (mm)	98.726	94.551	-6.775	-0.092
	(32.828)	(31.115)	(12.285)	
Average Temperature (°C)	13.944	14.327	-0.341	0.111
	(2.469)	(2.416)	(0.937)	
Wind Speed (m/s)	2.136	2.226	0.144	0.075
	(0.848)	(0.839)	(0.323)	
Observations	915	632	1,547	

Notes: The table reports the mean for the control municipalities (column 1), for treated municipalities (column 2), differences in mean (column 3) and standardized differences (column 4) of a series of geographical and climate outcomes. The analysis control for the population level in 1861 and an interaction term between the population level and the variable *Above4000*, that takes value 1 for municipalities with more than 4,000 inhabitants in the 1861 census. The sample is restricted to municipalities with population between 2,613 and 6,613. The outcomes are: elevation data, monthly climate data (solar radiation, minimum, mean, and maximum temperature, precipitation, water vapor pressure, wind speed, total precipitation), and an index that groups 19 "bioclimatic" variables. Elevation data is taken from the 1991 census, while the monthly climate data refers to October average from 1970-2000. We successfully matched all but 269 municipalities in our sample.

Figure C.1: Manipulation Testing Plot



Notes: The figure plots the manipulation test at the 4,000 inhabitants cutoff. P-value: 0.796.

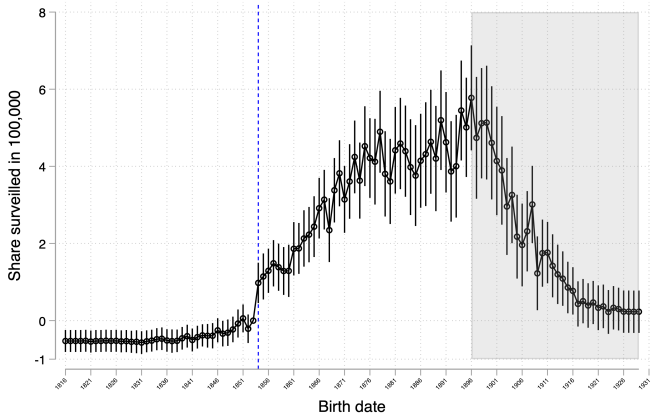
D Robustness to the Main Result

Table D.1: Effect on surveilled per 100,000 inhabitants, born after 1896

	(1) Share Surveilled	(2) Share Surveilled	(3) Share Surveilled
Post × Above4000 = 1	2.198***	2.073***	2.650***
	(0.260)	(0.253)	(0.244)
Pop1861 × Post	0.000545***	0.000526***	0.000792***
	(8.03e-05)	(7.96e-05)	(7.86e-05)
Pop1861 × Above4000	3.75e-06**		
	(1.56e-06)		
Pop1861 × Post × Above4000	-0.000527***	-0.000506***	-0.000773***
	(8.13e-05)	(8.06e-05)	(7.94e-05)
Observations	555,336	555,336	555,336
Birth FE	Y	Y	Y
Municipality FE	N	Y	Y
Reign x Birth FE	N	N	Y
Mean DV	0.0824	0.0824	0.0824

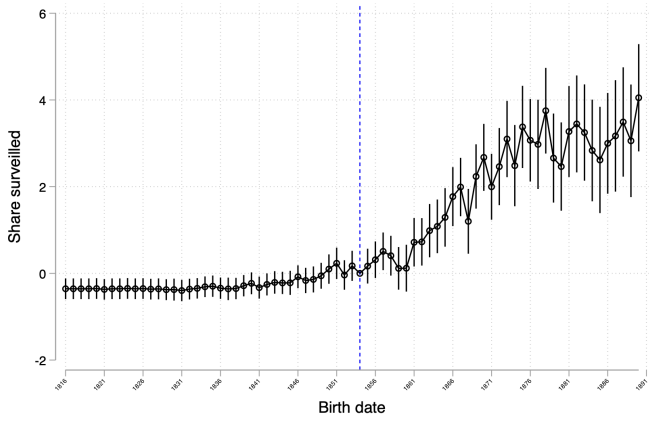
Notes: DDRD results as in Equation 1 of the effect of the education reform on the number of surveilled per 100,000 inhabitants in a municipality-cohort. The sample includes treated cohorts born from 1896 onwards and control cohorts born pre-reform. We report the mean of the dependent variable for cities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

Figure D.1: Effect of the Education Reform during and after the Discontinuity Applies on Surveilled in 100,000 (Extended Event Study)



Notes: The figure plots the effect of the education reform on the number of surveilled per 100,000 inhabitants. We use the same specification in Figure 4 for each year of birth from 1816 to 1929 (rather than stopping at cohort 1896). From cohort 1896, the education reform applies to all municipalities and the discontinuity stops applying. The regression includes municipality, date of birth, and reign times date of birth fixed effects. The standard errors are clustered at the municipality level.

Figure D.2: Effect on surveilled per 100,000 inhabitants (DiD)



Notes: DD results as in Equation 1 of the effect of the law on the number of surveilled per 100,000 inhabitants by year. The absorbed category is 1853 ($t-1$ before treatment). The variable identifying the territorial variation takes value 1 for all cities that register more than 4,000 inhabitants in 1861 for all the period. The regression includes municipality, date of birth and Reign times year fixed effects. The standard errors are clustered at the municipality level.

Table D.2: Alternative Dependent Variables

	(1)	(2)	(3)
		Surveilled	
	Any	N	Ln
Post × Above4000	0.276***	0.00788	1.948***
	(0.0145)	(0.200)	(0.132)
Observations	617,040	617,040	617,040
Municipality FE	Y	Y	Y
Birth FE	Y	Y	Y
Reign x Birth FE	Y	Y	Y
Mean DV	0.00100	0.00100	-6.898

Notes: Robustness to DDRD results as in Equation 1 where we vary the definition of the dependent variable. In column 1 we consider a dummy equal to one if a city has at least one surveilled born in that year; in column 2, we consider the absolute number of surveilled; in column 3, we consider the natural logarithm of the number of surveilled.

Table D.3: Robustness: higher order polynomial

	(1)	(2)	(3)
	Share Surveilled in 100,000		
Post × Above4000	3.515***	3.173***	2.466***
	(0.318)	(0.425)	(0.488)
			(0)
Observations	617,040	617,040	617,040
Municipality FE	Y	Y	Y
Birth FE	Y	Y	Y
Reign x Birth FE	Y	Y	Y
polynomial	1	2	3
Mean DV	0.0824	0.0824	0.0824

Notes: Robustness to DDRD results as in Equation 1 for higher order polynomials. Column (1) includes the first order polynomial of the difference-in-discontinuities variables; column (2) the second order polynomial; column (3) the third order polynomial. The table reports the mean of the dependent variable in municipalities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

Table D.4: Robustness: functional forms

	(1)	(2)	(3)	(4)
	Share Surveilled in 100,000			
	Control	Poisson	Neg Bin	Staggered Adoption
Post × Above4000	3.374***	1.251***	3.615***	3.275***
	(0.312)	(0.0123)	(0.020)	(0.284)
Observations	617,040	425,520	425,520	701,600
Municipality FE	Y	Y	Y	Y
Birth FE	Y	Y	Y	Y
Reign x Birth FE	Y	-	-	Y
Mean DV	0.0820	0.0820	0.0820	0.0920

Notes: Robustness to DDRD results as in Equation 1. Column (1) includes a control for the number of individuals born in the same city who begin surveillance 29 years after the cohort's birth year. Column (2) considers a different functional form – a poisson specification. Column (3) accounts for regions that were annexed to the Kingdom of Italy after 1861 and that implemented the Casati Law following their annexation. The staggered adoption is as follows: Friuli Venezia Giulia, Veneto, and the province of Mantova enacted the law in 1866 when they were annexed to the Kingdom of Italy; Lazio followed in 1870. The rest of Italy adopted the reform in 1861. We exclude Trieste, which was annexed in 1954. The table reports the mean of the dependent variable in municipalities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

Table D.5: Effect of the education reform, year of birth control ("Double" DDRD)

	(1)	(2)	(3)
	Share Surveilled in 100,000		
Post × Above4000	3.729***	3.605***	3.514***
	(0.359)	(0.350)	(0.318)
Observations	617,040	617,040	617,040
Birth FE	Y	Y	Y
Municipality FE	N	Y	Y
Reign x Birth FE	N	N	Y
Mean DV	0.0824	0.0824	0.0824

Notes: The table shows a "double" Difference-in-Discontinuity design in which we replicate Equation 1 with additional controls for the flexible impact of treatment across cohorts, allowing to isolate the effect of a possible jump caused by cohorts being exposed longer to the reform.

Table D.6: Effect of the Education Reform in a DiD setting

	(1)	(2)	(3)
	Share Surveilled in 100,000		
Post × Above4000	3.128***	3.013***	2.435***
	(0.258)	(0.251)	(0.216)
Observations	617,040	617,040	617,040
Birth FE	Y	Y	Y
Municipality FE	N	Y	Y
Reign x Birth FE	N	N	Y
Mean DV	0.0824	0.0824	0.0824

Notes: The table shows a difference-in-difference results of the effect of the education reform on the number of surveilled per 100,000 inhabitants in a municipality-cohort. The variable *Post* takes value 1 for cohort born from 1854 onwards (seven years old or younger when the reform was enacted). The variable *Above*4000 takes value 1 for municipalities with more than 4,000 inhabitants in the 1861 census. We do not control flexibly for population around the cutoff as in the main specification in Equation 1 We report the mean of the dependent variable for cities with population below 4,000 in 1861. Standard errors are clustered at the municipality level.

Table D.7: Effect of the Education Reform on Surveillance, by Democracy and Dictatorship

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Share Surveilled before 1922				Share Surveilled after 1922			
	Overall	Working Class	Middle Class	Upper Class	Overall	Working Class	Middle Class	Upper Class
Post × Above4000	1.831***	1.366***	0.234***	0.149***	1.683***	1.125***	0.240***	0.213***
	(0.199)	(0.153)	(0.0303)	(0.0306)	(0.164)	(0.130)	(0.0306)	(0.0259)
Observations	617,040	617,040	617,040	617,040	470,493	470,493	470,493	470,493
Birth FE	Y	Y	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y	Y	Y	Y	Y
Mean DV	0.0744	0.0474	0.0107	0.0115	0.0159	0.00870	0.00180	0.00150

Notes: DDRD results as in Equation 1 subsetting by start of surveillance. Columns (1)-(3) focus on individuals who begin surveillance before 1922; columns (4)-(6) on individuals who begin surveillance after 1922 –included. We report the mean of the dependent variable for cities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

Table D.8: Effect of the Education Reform on Surveillance, by ideology

	(1)	(2)	(3)	(4)
	Share Surveilled in 100,000			
	Anarchists	Anti-Fascists	Communists and Socialists	Others
Post × Above4000	0.836***	0.418***	2.065***	0.211***
	(0.148)	(0.0568)	(0.161)	(0.0523)
Observations	617,040	617,040	617,040	617,040
Birth FE	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y
Mean DV	0.0285	0.00350	0.0408	0.0111
Number of individuals	20522	13342	39692	6426
Δ over mean	29.33	119.5	50.60	18.99

Notes: DDRD results as in Equation 1 subsetting by ideology. We report the mean of the dependent variable for cities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

Table D.9: Placebo population discontinuity (below threshold)

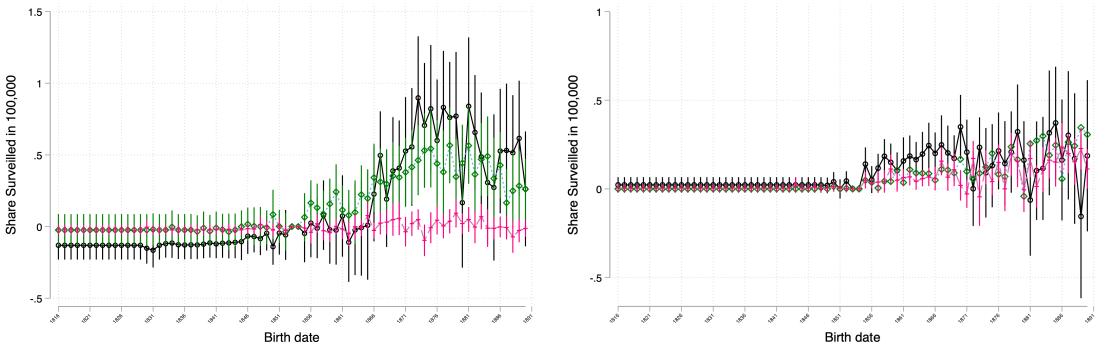
	(1)	(2)	(3)
	Above 1000	Above 2000	Above 3000
Post × Above4000	0.841	1.266	-0.937
	(0.566)	(1.280)	(3.627)
Observations	565,506	565,506	565,506
City FE	Y	Y	Y
Birth FE	Y	Y	Y
Reign x Birth FE	Y	Y	Y
Mean DV	.1066	.1001	.099

Notes: DDRD results as in Equation 1 but in municipalities with alternative (placebo) population cutoffs. In this table, we focus on municipalities with population level below 4,000 inhabitants in 1861 (all control units). All regression controls for the difference-in-discontinuities variables: Population × treat, Population × post, Population × treat × post. Standard errors are clustered at the municipal level.

Figure D.3: Effect of the Education Reform on Surveillance, by ideology and social class (Event study)

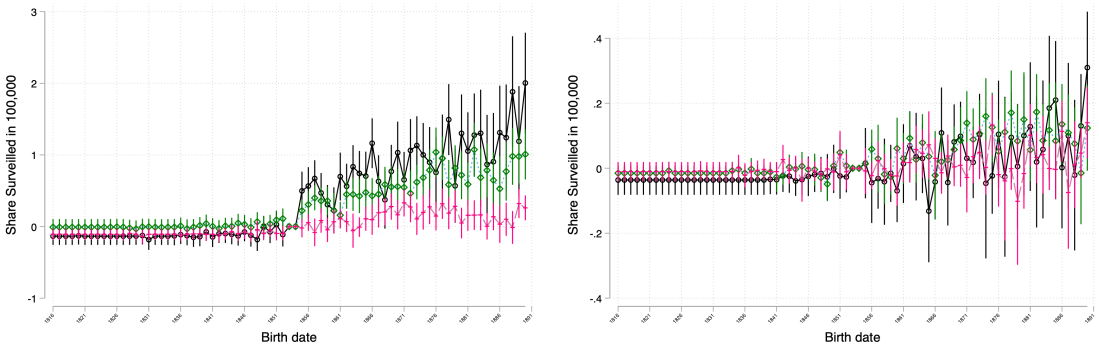
(a) Anarchists

(b) Anti-Fascists



(c) Communists and Socialists

(d) Others



Notes: DDRD results as in Equation 1 where the dependent variable includes only working (black), middle (green), and upper class (pink) subsetting by ideology. The absorbed category is 1853.

Table D.10: Placebo population discontinuity (above threshold)

	(1) Correct Cutoff	(2) Above 6000	(3) Above 8000	(4) Above 10000	(5) Above 12000	(6) Above 14000	(7) Above 16000	(8) Above 18000	(9) Above 20000	(10) Above 25000	(11) Above 30000
Post × Above10000	2.708***	0.230	-3.294	1.305	7.074	-18.74	-24.29	47.90	22.60	-48.37	-28.35
	(0.840)	(2.867)	(3.295)	(4.592)	(12.11)	(16.86)	(19.54)	(35.67)	(21.42)	(42.99)	(27.14)
Observations	593,520	83,920	83,920	58,960	31,920	21,518	14,558	10,238	7,920	2,880	1,877
Birth FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Window	0-10k	4k-14k	4k-14k	5k-15k	7k-17k	9k-19k	11k-21k	13k-23k	15k-25k	20k-30k	25k-25k

Notes: DDRD results as in Equation 1 but in municipalities with alternative (placebo) population cutoffs. In all regressions, we consider a window of 10,000 inhabitants around the cutoff. For the correct cutoff (4,000 inhabitants), we use observations from 0 to 10,000 inhabitants. For placebo cutoffs, we consider 10,000 inhabitants windows around the placebo cutoffs, but including only treated unitss (e.g., municipalities above 4,000 inhabitants). All regression controls for the difference-in-discontinuities variables: Population × treat, Population × post, Population × treat × post. Standard errors are clustered at the municipal level.

Table D.11: Placebo Cohorts

	(1)	(2)	(3)	(4)
Post × Above4000	-1.748***	-1.891***	-1.847***	-1.642***
	(0.168)	(0.180)	(0.180)	(0.166)
Observations	578,475	578,475	578,475	578,475
Birth FE	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y
Year	1810	1820	1830	1840

Notes: DDRD results as in Equation 1 but for alternative (placebo) cohorts that are not treated. We consider cohorts between a certain threshold and that threshold plus 10. The regression controls for the difference-in-discontinuities variables: Population × treat, Population × post, Population × treat × post. Standard errors are clustered at the municipal level.

Table D.12: Heterogeneity by level of surveillance pre-reform

	(1)
	Share Surveilled in 100,000
Post × Above4000	2.892***
	(0.336)
Treatment × HighSurveillance	4.083***
	(0.788)
Observations	617,040
Birth FE	Y
Municipality FE	Y
Reign x Birth FE	Y
Mean DV	0.0824

Notes: Heterogeneity to DDRD results as in Equation 1 by pre-reform levels of surveillance. The dummy *HighSurveillance* takes value 1 if a municipality has a high level of pre-reform surveillance. A municipality is classified as having a high level of pre-reform surveillance if its average surveillance level before the reform is greater than or equal to the median pre-reform surveillance level. The table reports the mean share of surveilled in untreated municipalities before the reform. Standard errors are clustered at the municipality level.

Table D.13: Effect of the Education Reform on Surveillance, by geographical area

	(1)	(2)	(3)
	Share Surveilled in 100,000		
	North	Center	South
Post × Above4000	5.279***	8.502***	1.651***
	(0.521)	(1.094)	(0.243)
Observations	351,920	65,120	200,000
Birth FE	Y	Y	Y
Municipality FE	Y	Y	Y
Reign x Birth FE	N	N	N
Mean DV	0.0946	0.143	0.0389

Notes: Heterogeneity to DDRD results as in Equation 1 considering whether the presence of any (column 1) or the number of (column 2) postal offices in a municipality. The table reports the mean of the dependent variable in municipalities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

Table D.14: Effect of the Education Reform on Surveillance, by presence of post offices

	(1)	(2)
	Share Surveilled in 100,000	
Post × Above4000	2.855***	2.866***
	(0.359)	(0.335)
Post× Above4000 × Any Post Office	1.203***	
	(0.427)	
Post× Above4000 × N Post Offices		1.185***
		(0.441)
Observations	617,040	617,040
Birth FE	Y	Y
Municipality FE	Y	Y
Reign x Birth FE	Y	Y
Mean DV	0.0760	0.0760

Notes: Heterogeneity to DDRD results as in Equation 1 subsetting by the presence of a post office. The dummy *Any Post Office* takes value 1 if there is at least one working post office in that municipality in 1861. The variable *N Post Offices* is a continuous variable that corresponds to the number of working post offices in that municipality. The table reports the mean of the dependent variable in municipalities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

E Additional Analyses on Mechanism

Table E.1: Effect on Surveilled by Class (Dummy)

	(1)	(2)	(3)
	Any Surveilled		
	Working class	Middle class	Upper class
Post × Above4000	0.211***	0.0301***	0.0278***
	(0.0141)	(0.0103)	(0.00964)
Observations	617,040	617,040	617,040
Birth FE	Y	Y	Y
Municipality FE	Y	Y	Y
Reign x Birth FE	Y	Y	Y
Mean DV	0.0122	0.00360	0.00370
Δ over mean	17.33	8.371	7.523

Notes: Difference-in-discontinuity results (Equation 1) of the effect of the education reform on a dummy for whether any individual from that social class is surveilled in a municipality-cohort. The table report the mean of the dependent variable for cities with population below 4,000 in 1861 before the reform. The regressions include year of birth, municipality and reign times year fixed effects. Standard errors are clustered at the municipal level.

Table E.2: Effect on Political Confinement

	(1)	(2)	(3)	(4)
	Share Sentenced to Confinment in 100,000			
	Overall	Working Class	Middle Class	Upper Class
Post × Above4000	0.867***	0.486***	0.272***	0.0766***
	(0.0969)	(0.0711)	(0.0328)	(0.0167)
Observations	617,040	617,040	617,040	617,040
Birth FE	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y
Mean DV	0.0009	0.0004	0.0004	0.0001
Δ over mean	963.2	1214	681.1	766.5

Notes: The table displays the results for 1 where the dependent variable is the number of individuals sentenced to political confinement (*confino*) in 100,000 inhabitants by class. We report the mean of the dependent variable for cities with population below 4,000 in 1861 before the reform. The regression includes the controls for Population in 1861 x post and Pop in 1861 x post x treat. Standard errors are clustered at the municipality level.

Table E.3: Effect on Duration of Surveillance

	(1)	(2)	(3)	(4)
	Share of the year subject to surveillance			
	Overall	Working Class	Middle Class	Upper Class
Post × Treat	5.176***	3.075***	1.744***	0.586***
	(0.317)	(0.274)	(0.298)	(0.189)
Observations	617,040	617,040	617,040	617,040
Birth FE	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y
Mean DV	0.0379	0.0166	0.0126	0.00670
Mean days surveilled	13.83	6.059	4.599	2.446
Δ over mean	136.6	185.3	138.4	87.40

Notes: The table display the effect of the reform on the duration of surveillance for the entire sample (column 1) and by class (columns 2-4). The dependent variable is the share of the year a municipality-cohort spends under surveillance. We translate this share in average number of days in the bottom of the table. The mean is calculated for cities with population below 4,000 in 1861, before the reform. $\Delta overmean$ shows the effect size by dividing the coefficient over the mean of the DV at baseline. Standard errors are clustered at the municipal level.

Table E.4: Effect on quantity of information collected

	(1)	(2)	(3)	(4)
	Share Surveilled with biographical file in 100,000			
	Overall	Working Class	Middle Class	Upper Class
Post × Above4000	1.361***	0.667***	0.500***	0.157***
	(0.134)	(0.0818)	(0.0495)	(0.0268)
Observations	617,040	617,040	617,040	617,040
Municipality FE	Y	Y	Y	Y
Birth FE	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y
Mean DV	0.0517	0.0227	0.0177	0.00940
Δ over mean	26.33	29.38	28.26	16.66

Notes: The table reports the results for 1 where the dependent variable is the number of individuals surveilled with biographical information in 100,000 inhabitants. Column (1) shows the results for the whole sample; columns (2)-(4) by class. The variable identifying the territorial variation takes value 1 for all cities that register more than 4,000 inhabitants in 1861 for all the period. The regression includes the controls for Population in 1861 x post and Pop in 1861 x post x treat. The regressions include birth year, municipality and reign time year fixed effects. Standard errors are clustered at the municipality level.

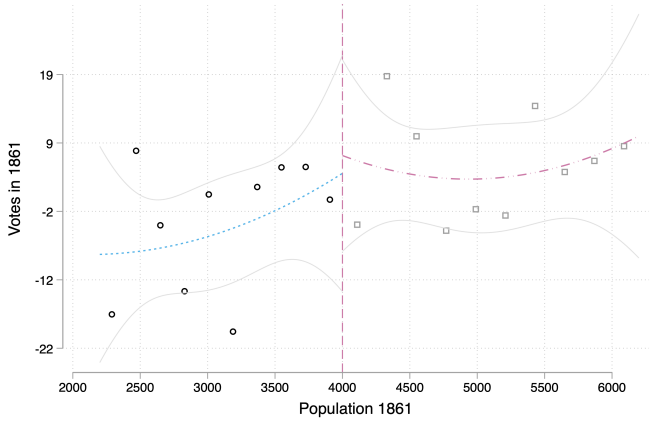
F Additional Analyses on Individual Political Activation

Table F.1: Effect on voter mobilization (DDRD)

	(1) Number of Votes
Post × Above 4000	23.60
	(123.7)
Observations	61,648
Municipality FE	Y
Election Year FE	Y
Reign x Election Year FE	Y
Mean DV	162.7

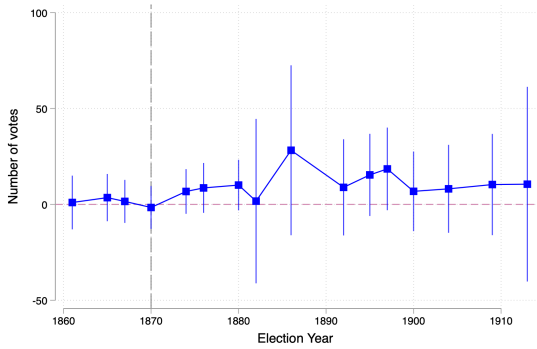
Notes: Difference-in-Discontinuity results of the effect of the education reform on the number of votes. Individuals who were 7 years or younger at the time of the reform voted for the first time in 1880, allowing us to consider four election years as pre-periods before the start of treatment ($Post$).

Figure F.1: Effect of the 4,000 cutoff on mobilization in 1861 (RDD)



Notes: Regression discontinuity Calonico, Cattaneo and Titiunik (2015) showing the effect of being above the 4,000 population discontinuity on the absolute number of votes cast in 1861. The figure provides evidence in favor of the continuity of potential outcomes assumption by showing the lack of effects before the first cohort treated by the education reform started to vote.

Figure F.2: Effect on voter mobilization by election year (RDD)



Notes: Regression discontinuities Calonico, Cattaneo and Titiunik (2015) showing the effect of being above the 4,000 population discontinuity on the absolute number of votes cast in each election year. Individuals who were 7 years or younger at the time of the reform voted for the first time in 1880. Standard errors are clustered at the municipal level.

Table F.2: Effect on notable individuals from Wikidata

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		Non Politics		Politics	
	Share	Total	Share	Total	Share	Total
Post × Above4000	-0.294*	-0.0869**	-0.170**	0.0134***	-0.236***	0.00345
	(0.159)	(0.0361)	(0.0790)	(0.00256)	(0.0860)	(0.00398)
Observations	328,480	328,480	328,480	328,480	328,480	328,480
Municipality FE	Y	Y	Y	Y	Y	Y
Birth FE	Y	Y	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y	Y	Y
Mean DV	0.669	0.0102	0.231	0.00410	0.259	0.00450
Δ over mean	-0.439	-8.517	-0.737	3.274	-0.910	0.766

Notes: Treatment effect (being 7 years old or younger in 1861) on the number of notable individuals included in Wikidata per 100,000 inhabitants. Columns (1) and (4) report the effect on all notable individuals. Columns (2) and (5) focus on non-political occupations, while columns (3) and (6) consider political occupations (see Section A for variable construction). The regressions control for year of birth, municipality, and reign-by-year of birth fixed effects. Standard errors are clustered at the municipality level.

Table F.3: Effect on notable individuals from Wikidata by surveillance level

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		Non Politics		Politics	
	Share	Total	Share	Total	Share	Total
Post × Above4000	-0.301*	-0.0907**	-0.176**	0.0140***	-0.241***	0.00363
	(0.160)	(0.0386)	(0.0791)	(0.00267)	(0.0861)	(0.00421)
Post × Above4000 × HighSurveillance	0.0470	0.0254	0.0377	-0.00392	0.0360	-0.00126
Observations	328,480	328,480	328,480	328,480	328,480	328,480
Birth FE	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y
Reign x Birth FE	Y	Y	Y	Y	Y	Y
Mean DV	0.669	0.0102	0.231	0.00410	0.259	0.00450

Notes: Treatment effect (being 7 years old or younger in 1861) on the number of notable individuals included in Wikidata per 100,000 inhabitants by surveillance level pre-reform. Columns (1) and (4) report the effect on all notable individuals. Columns (2) and (5) focus on non-political occupations, while columns (3) and (6) consider political occupations (see Section A for variable construction). The dummy *HighSurveillance* takes value 1 if the median number of individuals surveilled from untreated cohorts in a municipality is larger than the median at the national level. The regressions control for year of birth, municipality, and reign-by-year of birth fixed effects. Standard errors are clustered at the municipality level.

Table F.4: Effect on Protests

	(1)	(2)	(3)	(4)	(5)	(6)
	N Protests in 100,000		Winsorized		Ln Protests	
RD Estimate	-106.0	-120.0*	-13.37	-15.64	0.0798	0.0495
	(65.50)	(71.35)	(14.94)	(15.98)	(0.189)	(0.198)
Observations	291,609	291,609	291,609	291,609	294,872	294,872
Bandwidth	1450.32	2448.45	1549.55	2803.30	2108.38	4050.57
N Left	256055	256055	256055	256055	259318	259318
N Right	35554	35554	35554	35554	35554	35554
Robust P-val	0.160	0.046	0.386	0.186	0.794	0.772
Mean DV	109.299	109.299	20.827	20.827	-6.462	-6.462
Polynomial	1	2	1	2	1	2

Notes: The table shows the effect of the average number of protests in cities with populations above the 4,000 cutoff. The table reports the average dependent variable in the sample. Standard errors are clustered at the municipal level.

Table F.5: Effect on partisans activity

	(1)	(2)	(3)
		Partisans	
	Share	Total	Dummy
Post × Above4000	-0.0973	-0.0723***	0.0413***
	(0.151)	(0.0258)	(0.00477)
Observations	617,040	617,040	617,040
Birth FE	Y	Y	Y
Municipality FE	Y	Y	Y
Reign x Birth FE	Y	Y	Y
Mean DV	0.00270	0.00004	0.00004

Notes: DDRD results as in Equation 1 but using as dependent variable partisans rather than surveilled. We consider the share of partisans in 100,000, as in the main surveillance specification (column 1), the total number of partisans (column 2), and a dummy for whether any partisan was recorded in that municipality-cohort. Standard errors are clustered at the municipality level.

G Alternative Interpretations for the Main Result

Table G.1: Placebo Reform - More Bureaucrats but Not More Schools

	(1)	(2)	(3)
Post × Above5000	8.896	8.894	8.894
	(5.454)	(5.454)	(5.454)
Observations	95,827	95,827	95,827
City FE	N	Y	Y
Year FE	Y	Y	Y
Regno x Year FE	N	N	Y
Mean DV	0	0	0

Notes: The table reports the effect of the Crispi reform enacted in 1890 in municipalities above 5,000 inhabitants on the share of surveilled in 100,000. The database is at the municipality and year of start of surveillance level in municipalities with population below 5,000 before 1890. Municipalities with less than 4,000 inhabitants in 1861 are not included in the analysis. The table report the mean share of surveilled in municipalities with population level below 5,000 before the reform. Standard errors are clustered at the municipality level. See Reforms section for details on the Law.

Table G.2: Effect of the Education Reform on non-migrants

	(1)	(2)	(3)
	Share Surveilled in 100,000		
Post × Above4000	2.100***	2.036***	2.075***
	(0.218)	(0.212)	(0.194)
Observations	617,040	617,040	617,040
Birth FE	Y	Y	Y
Municipality FE	N	Y	Y
Reign x Birth FE	N	N	Y
Mean DV	0.0516	0.0516	0.0516

Notes: Difference-in-discontinuity results of the effect of the education reform on the number of surveilled per 100,000 inhabitants in a municipality-cohort (Equation 1). The analysis include only on non-migrants, individuals for which the registered birth and residence locations are the same. We report the mean of the dependent variable for cities with population below 4,000 in 1861 before the reform. Standard errors are clustered at the municipality level.

Table G.3: Effect of the Education Reform by different professions

	(1)	(2)	(3)	(4)	(5)
	N Teachers	Share Teachers	Share Shop Owners	Share Lawyers	Share Tailors
Post × Above4000	-0.000430	0.00480***	0.00403***	0.00704***	0.00807***
	(0.00258)	(0.000474)	(0.000371)	(0.000560)	(0.000523)
Observations	617,040	617,040	617,040	617,040	617,040
Birth FE	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y
Mean DV	5.00e-05	5.00e-05	2.00e-05	3.00e-05	4.00e-05
Δ over mean	-8.602	96.05	201.7	234.5	201.7

Notes: Difference-in-discontinuity results of the effect of the education reform on the number of individuals with a given profession surveilled in a municipality-cohort (Equation 1). We consider the absolute number of teachers in column 1, and in the following columns the share of teachers, and of other professions surveilled at similar rates pre-reform. $MeanDV$ is the dependent variable mean for cities below 4,000 inhabitants before the reform. $\Delta overmean$ shows the effect size by dividing the coefficient over the mean of the DV at baseline. Standard errors are clustered at the municipality level.