Local scandals of corruption and turnout. Evidence from Italian elections.

Ruben Piazzesi

Thesis advisor: Prof. Facundo Piguillem

Einaudi Institute for Economics and Finance (EIEF)
LUISS Guido Carli

Abstract

This paper aims to study the effects of exposure to local corruption on electoral turnout. Focusing on a sample of 370 Italian municipalities in the period 2006–2020, I generate a local measure of exposure to political corruption, screening newspaper articles of the leading Italian press agency. I use electoral data at the municipal level for three types of elections: municipal, regional, and national. By estimating two fixed-effects models, I find that political corruption and turnout are either uncorrelated or negatively correlated at the local level. Therefore, I reject the hypothesis that political corruption increases turnout, supported by scholars in the previous literature. I then test the hypothesis that exposure to local corruption might affect turnout in local and national elections differently. I do not find suggestive evidence of such a heterogeneous effect. Finally, I test whether corruption and turnout correlate differently in provinces and not provinces. I do not find conclusive results, leaving the question open for future studies.

1 Introduction

Corruption is a broad phenomenon with sizeable and complex effects on society. Empirical evidence shows that corruption negatively affects voters' trust towards public institutions (Clausen et al., 2011), as well as their confidence in the governments' capability (Caillier, 2010). In this context, elections become a crucial vehicle of transmission. Through the elections, the voters can directly express their discontent and bring down corrupt governments. Empirical studies have widely documented the existence of a *disciplinatory channel* from the side of voting choices. For instance, in corrupted localities, incumbents face a lower probability of being re-elected in municipal elections (Ferraz and Finan, 2008) (Costas-Pérez et al., 2012), and after the elections, the subsequent administrators tend to behave more honestly (Avis et al., 2018). While most empirical studies focus on how voters sanction incumbents upon learning about their misbehavior while in office, fewer observational studies document the effect of corruption on electoral turnout. Nonetheless, a better understanding of such a relationship is desirable for several reasons. First, the electoral outcomes are the result of two consecutive decisions. Before deciding the candidate to support at the elections, voters face a participation decision, whether to vote or abstain. Not taking into account the first step of the voting decision might lead to bias estimations of the effect of corruption on other electoral results (Sundström and Stockemer, 2013), (Stockemer, 2013). Second, voter turnout is one of the most important indicators of the level of the democratic legitimacy of a country. There is evidence that corruption is especially disruptive in democracies because it undermines the basic principles of a free state (Rock, 2009). Whether corruption also dissuades citizens from turning out in democratic elections is particularly important because such lower participation may exacerbate legitimacy problems. Hence, the main question that I aim to address in this paper is how local scandals of political corruption affect turnout. To do so, I use electoral data from the Italian elections in the period 2006-2020, collected at the municipal level.

The effect of corruption on electoral turnout is *ex-ante* ambiguous, as the previous literature has found mixed results. Among the few empirical works that have focused on the issue, three opposite claims are made. On the one hand, claims and empirical evidence are put forward saying that corruption decreases turnout. From a theoretical point of view, the presence of

corruption scandals affects voter's rationality, increasing the voting cost (Heidenheimer Arnold et al., 1989), (Matsusaka, 1995). This, consistently with the Downsian model (Downs et al., 1957) of the rational voter, whose decision to participate in elections hinges on a comparison of the benefits and costs of voting, would lead to a reduction in electoral turnout. From an empirical perspective, evidence supporting the idea of a negative relationship is provided by Chong et al. (2015), who focus on local incumbents in Mexico and find that the dissemination of information on politicians' malfeasance lowers voter turnout. Additionally, Costas-Pérez (2013) finds similar results focusing on the case of Spain. The typical argument links political corruption to mistrust towards politicians. If the exposure of corruption causes voters to lose trust in governments, they could start to believe that voting will not benefit them, making abstention more likely. Exposing voters to information about political corruption could reduce their confidence in the responsiveness of electoral institutions and public officials (Ansolabehere et al., 1994), which in turns may lead to cynicism, distrust towards the political administration, and voter apathy (Kostadinova, 2009).

The counter-argument is that political corruption increases turnout. One reason offered in support of this hypothesis is that the misuse of the public office increases the expected returns of staying in office (for instance, through bribe-taking), leading the political parties to put more effort into their campaigns, mobilizing voters, and increasing turnout. Studies from the US, focusing on county-level elections (Lacombe et al., 2016) and gubernatorial elections (Escaleras et al., 2012) find empirical evidence supporting this hypothesis. Additionally, a different argument for why the presence of corruption should increase turnout is that voters might attempt, through higher participation at the ballot, to bring down corrupt governments (Bratton et al., 2005).

Finally, the third group of scholars either do not find any relationship between corruption scandals and voter turnout (Stockemer, 2013), (Persson et al., 2013), or find mixed results. For instance, Kostadinova (2009) shows that while in the short term corruption can stimulate voter turnout to bring down corrupt governments, in the long run, turnout could eventually fall if corruption is persistent. Additionally, Costas-Pérez (2013) shows that different types of corruption may have heterogeneous effect on voters' reaction.

Hence, neither theoretical nor empirical studies have arrived at unified conclusions on the re-

lationship between corruption and voter turnout. This paper contributes to this stream of literature by providing evidence from the Italian elections in the period 2006-2020. Italy, in fact, constitutes an important example to be studied in terms of political corruption for different reasons. First, it represents an important outlier in the international panorama, as it is considered one of the most corrupted nations among the developed countries, in particular in the EU area (tra, 2020). Second, after the large corruption scandal that occurred at the beginning of the 90s, *Mani Pulite*, that has deeply modified Italian politics (Vannucci, 2009), the phenomenon of political corruption has become broader in the last two decades (Vannucci, 2010), especially at the municipal level (Anticorruzione, 2019).

Few empirical works have explored the connection between political corruption and turnout in the Italian case. Giommoni (2021) focuses on the *indirect effect* of political corruption that occurred in local capitals on turnout in the other cities, finding a negative relationship. Additionally, Daniele et al. (2020) document a negative long-run effect on turnout after the national scandal of *Mani Pulite*. In this paper, I contribute to this strand of literature by studying the *direct impact* of corruption scandals occurring at the municipal level. Importantly, I exploit turnout data for both local and national elections.

The paper is also related to the media economics literature, in particular the literature focusing on media reporting of corruption scandals (Giommoni, 2021), (Le Moglie and Turati, 2019), (Rizzica and Tonello, 2015). In order to measure political corruption, I screened newspaper articles of the main Italian press agency, Agenzia Nazionale Stampa Associata (*ANSA*), which cite local politicians involved in corruption cases. Among the paper's contributions is the creation of a novel dataset of corruption scandals involving politicians in office in a sample of 370 Italian municipalities in the period of 2006-2020.

In order to detect the correlation between corruption scandals and turnout, the baseline specification is a model with city fixed effects. Overall, I find a negative yet not statistically significant correlation. The results suggest that most of the variability in the data is due to unobservable and time-invariant characteristics at the municipal level that determines both corruption and turnout. While the city fixed-effects model is the benchmark, I cannot exclude that the identification strategy might suffer from some limitations present in the data. Estimating a model with city fixed effects could be problematic, given the quite small variation present in the dataset of

corruption. I propose a model that controls for fixed effects by zone (North, South, Center of Italy), finding a negative and statistically significant correlation. Additionally, the measure of corruption might be endogenous. In particular, I cannot exclude that the negative correlation found in the data could be driven by reverse causality: the lower the political involvement, the lower the control on the incumbent's activity, that in turn fosters corruption by decreasing the political accountability (Lederman et al., 2001).

I then explore the hypothesis that the impact of local corruption on turnout may differ according to the type of election in the analysis. While it is pretty immediate that local scandals could influence turnout in the local elections, the relation is less clear when it comes to national elections. I find no heterogeneous effect between different types of elections. The result holds both when I estimate the model with city fixed-effects and with zone fixed-effects.

Finally, this paper is related to the strand of literature that focuses on the effect of municipal characteristics on turnout. Van Houwelingen (2017) finds that more populated cities tend to face lower electoral participation. Additionally, previous works document that voters' behavior at the poll might be influenced by sociopolitical characteristics, such as the administrative power of a municipality and its degree of urbanization (Dahl and Tufte, 1973), (Van Houwelingen, 2017), (Monroe, 1977). I exploit the variation related to the status of province capitals, usually assigned to the biggest and most important cities, to test the hypothesis that corruption might have a different impact on turnout according to the size and administrative importance of the municipality. I find no heterogeneous effect when estimating the model with city fixed-effects. Instead, the zone fixed-effects model evidences that an additional scandal is associated with a 1.6 percent decrease in turnout in cities that are not province capitals, while the effect is negligible in province capitals.

To sum up, the paper exploits two different identification strategies. The baseline model evidences no significant correlation between corruption and turnout at the local level. On the other hand, the zone fixed-effects model documents a negative and significant correlation, even if the results might be driven by unobservables and time-invariant characteristics at the municipal level. Overall, the main results of the paper exclude that political corruption increases turnout, a hypothesis supported by the previous literature.

The rest of the paper is organized as follows. Section 2 describes the institutional background.

Section 3 describes the data. Section 4 presents the empirical design. Section 5 shows the main results. Section 6 describes the identification issues and the data limitations. Section 7 concludes the paper.

2 Institutional Background

2.1 Local administrations The Italian territory is divided into three primary administrative levels: regions, provinces, and municipalities. Currently, there are 20 regions, 107 provinces, and 7904 municipalities (ISTAT).[1] Each region has its local capital, and each province has its reference city. For example, Rome is both the capital of the region Lazio and the capital of the province of Rome. Statistically, most Italian municipalities are small cities, with an average population of 7334 per city. Province capitals are typically the most important cities, with an average population of 88.000 inhabitants and a bigger administrative power.[2]
Each administrative level is characterized by specific competencies and has its local government. In this paper, I use a measure of local corruption that explicitly focuses on corruption scandals involving politicians in office in the municipal government. The municipal governments are in charge of the local management. It is composed of a mayor (*Sindaco*), an executive committee (*Giunta*) and a city council (*Consiglio Comunale*). The number of components of the municipal government varies according to the population of the city. Therefore, the number of municipal politicians is typically larger in provincial capitals. Today, Italian municipalities are important providers of public services, including waste management, kindergartens, police, sports facilities, roads, street lighting, and urban planning. In terms of local public finance, municipal revenues come from taxes, fees, capital transfers, sales of public assets, borrowing, and transfers, either from central or regional governments or from the European Union.
In Italy, three different types of elections occur, with a frequency of 5 years: national elections, regional elections, and municipal elections. The current electoral system for municipal elections uses a single ballot, with plurality rule, for cities with less than 15,000 inhabitants and runoffs

[1] In the last decade, many attempts at reforming the institution of the province have been proposed. Since 2011 the local government of the province is not anymore elected by the citizens. Furthermore, some provinces have been suppressed over time, and some others (typically the most populated provinces) have been assigned the status of *metropolitan city*. The number of provinces that are reported here includes both the suppressed provinces and the *metropolitan city*.

[2] This average only takes into account the population of province capitals that are not also Region Capitals. If we also include region capital, the average population is likely to be far higher.

for the others, as well as the direct election of the mayor. The regional elections use a single ballot electoral rule that attributes part of the council seats based on a proportional system (80 percent) and the others based on a majority rule as well as the direct election of the President of the region. Finally, the electoral system for national elections uses a majority bonus system for the assignment of the council seats for both the branches of the parliament, Camera and Senato. [3]

2.2 Corruption in Italy Corruption is a widespread phenomenon in Italy. The country represents a significant outlier in the international panorama, especially among the western democracies. This is documented, for instance, by the Transparency International's Corruption Perception Index (CPI), which has ranked Italy in position 52nd out of 180 countries in the 2020 ranking of corruption perception, and 20th out of 27 countries in the in EU panorama (tra, 2020).

The problem of corruption has become a significant public issue in Italy after the enormous national scandal of Clean Hands (*Mani pulite*) in 1992, which uncovered a vast corruption scheme involving political parties as well as economic and public institutions. Starting from this judicial prosecution, many other scandals have been discovered in the following years (Vannucci, 2010). Notably, during the 2000s, corruption started to spread also among local administrations (Anticorruzione, 2016).

The Italian government has exerted notable effort over the years to contrast the widespread diffusion of corruption. Among the central interventions, two laws against corruption have been approved by the Italian Parliament in the last decade (2012 and 2019). The aim was to contain the spread of corruption within public administrations through a series of new measures, such as introducing new transparency procedures for public procurement and increasing the penalties for corruption-related crimes. A crucial intervention was also the constitution of a national anti-corruption agency in 2014, Autorità Nazionale Anticorruzione *(ANAC)*, which aims at detecting and fighting corruption using surveillance and auditing authorities, as well as at providing reports about the level of corruption in Italy at the yearly base.

According to the 2019 report, corruption in Italy is particularly widespread within local admin-

[3]The electoral system for national elections has changed after the national elections in 2013. Among other modifications, the size of the majority premium has changed substantially.

istrations (Anticorruzione, 2019). In the period 2016-2019, more than 40 percent of the episodes of corruption were related to sectors under the jurisdiction of the municipal governments. In the same period, the phenomenon of political corruption has kept a primary role, with 25 percent of the cases of corruption involving a politician in office in the municipal governments. Given the widespread political corruption at any administrative level, Italy is an ideal candidate for exploring the correlation between corruption and turnout. Additionally, the data released by ANAC, which documents a pattern of diffusion of political corruption within municipalities, justify the choice of focusing on scandals related to municipal governments.

3 Data

3.1 Selection of the sample The first contribution of this paper is the creation of a measure of exposure to local corruption in the context of Italian municipalities. For the data collection, I relied on the leading national press agency in Italy, Agenzia Nazionale Stampa Associata (*ANSA*). It is a widely used source of information in the literature of media economics (Giommoni, 2021) since it offers several convenient features. First, it is a key institution in the Italian press market, and it is not likely to be prone to ideological or geographical biases. Second, it provides good coverage of local news, especially through the service *ANSA Regional News*.

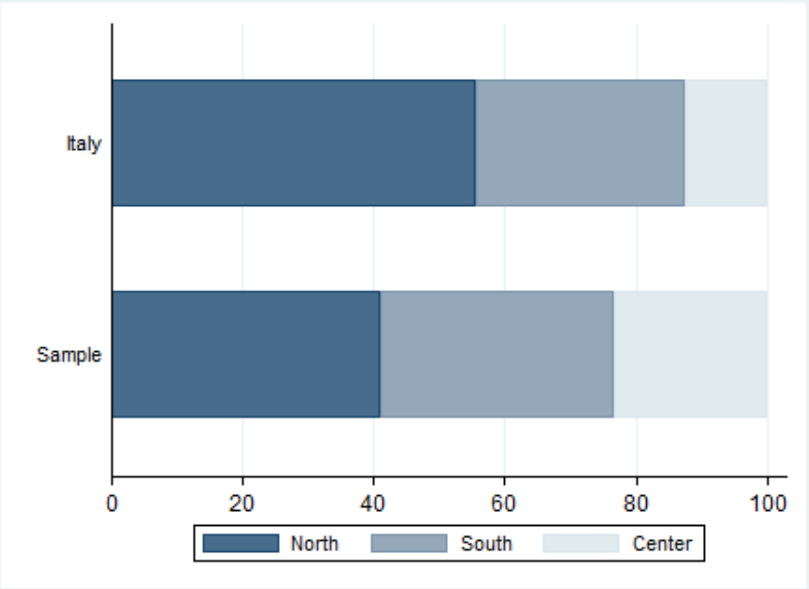
As discussed in section 2.1, most of the Italian municipalities are small and geographically dispersed. Concerning these cities, the coverage of ANSA is likely to be broad and imprecise. To deal with this issue, I decided to exclude from the sample in analysis the cities with fewer than 20.000 inhabitants. On the other hand, the opposite problem could arise when dealing with regional capitals. The share of news reported annually by ANSA for region capitals is much higher than any other city in Italy.[4] To avoid a massive overrepresentation of a small subsample of cities, I dropped from the sample all the regional capitals. Finally, I cut from the sample all the municipality located in four Italian regions (out of twenty) due to the unavailability of turnout data: Aosta Valley, Friuli-Venezia Giulia, Sicily, and Trentino-Alto Adige.

The final sample of cities is composed of 370 municipalities located in sixteen Italian regions. Figure 1 reports the geographic distribution of the cities, both in the sample and in the pool

[4]Figure 5 in Appendix I provides evidence of the substantial overrepresentation of regional capitals.

of all Italian municipalities, divided between North, South, and Center of Italy.[5] The North of Italy is slightly underrepresented in the sample, while the Center of Italy is overrepresented.

Figure 1 Geographic distribution of municipalities



Notes: The figure shows the geographic distribution of Italian municipalities. Three geographic areas are considered: North, South, and Center of Italy. The first row shows the proportion of municipalities belonging to each area among all Italian municipalities. The second row reports the same distribution, considering only the municipalities belonging to the sample in the analysis.

3.2 Data about political corruption To construct the measure of corruption, I downloaded articles from ANSA, using the digital library *Factiva*.[6] The period of the analysis is 2006-2020. I focused on this period because generally Factiva's coverage starts to be good after 2006.[7] Using an automatic keyword-based procedure, I then screened the articles' text to identify the articles about corruption scandals involving a local politician in office in the municipal government. The screening is based on three main criteria: a) the presence of the name and surname of a local politician in office in the city where the article was geolocalised, in the period when the article was written; b) the presence of at least one political office title in the municipal government (e.g., *mayor*, *assessor*, *councilor*); c) the presence of at least one

[5] Table 6 of Appendix I describes the composition of the geographic areas.

[6] Factiva is a business information and research tool owned by Dow Jones Company. Factiva aggregates content from licensed and accessible sources and provides organizations with search, alerting, dissemination, and other information management capabilities. Factiva products provide access to more than 32,000 sources (such as newspapers, journals, magazines, television and radio transcripts, photos) from nearly every country worldwide in 28 languages, including more than 600 continuously updated newswires.

[7] Figure 6 of Appendix I reports the evolution of Factiva's coverage over time.

keyword related to political corruption (e.g., *abuse of office*, *bribe*) in the same sentence as the surname of the local politician or the political office title. In Appendix II, I provide a detailed explanation of the screening strategy.

The procedure is fairly accurate in selecting articles of political corruption at the municipal level. Moreover, in order to exclude the presence of *false positive* that could potentially bias the estimations, I manually checked the sample of selected articles.[8] This procedure allows for the identification and geolocalization of 1477 articles about political corruption. In section 6 I will discuss some limitations of the corruption data that might threaten the identification strategy.

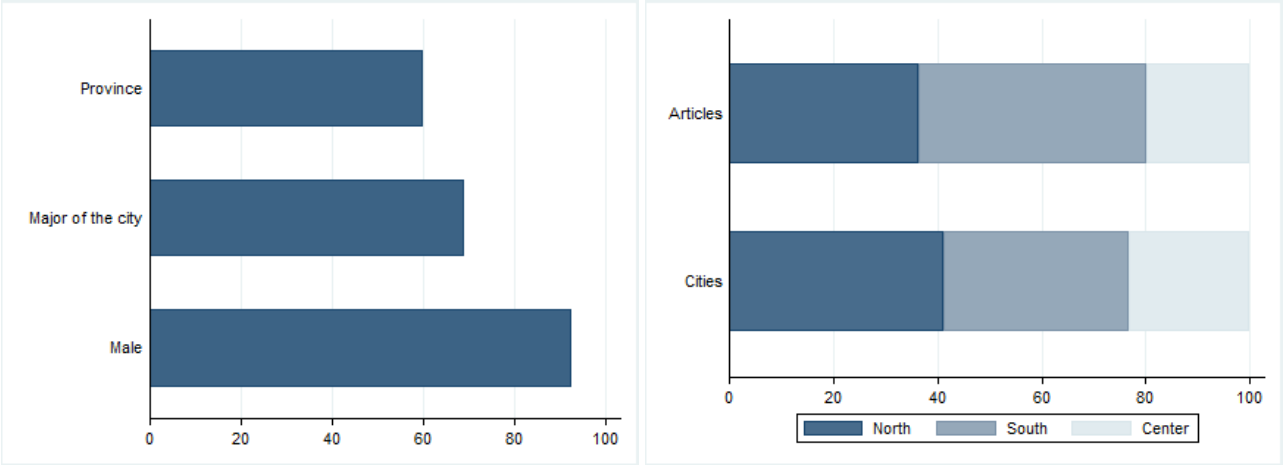
Figure 2.a reports a description of the final sample of articles. Almost 60 percent of the sample refers to scandals involving local politicians in office in provincial capitals. Such a disproportion might be due to different reasons. On the one hand, ANSA has better coverage of provincial capitals, the biggest and most important cities. On the other hand, one cannot exclude that provinces' politicians may be systematically more corrupted. Finally, as discussed in section 2.1, provinces count a larger number of politicians, a feature that is likely to increase mechanically the number of cases of corruption detected. Still, the share of articles geolocalized in not provinces is consistent, around 40 percent. This ensures a good representation of the whole sample of cities. Additionally, almost 70 percent of the scandals involve the city's mayor, and more than 90 percent of the articles involve male politicians.

Graph 2.b reports the geographic distribution of the articles in comparison with the distribution of the sample of cities in analysis. On average, politicians in office in the South of Italy are by far the most corrupted.

Finally, figure 3 describes the distribution over time of the screened articles at monthly frequency. The graph reports the evolution of a *corruption index* that takes into account possible variations in the coverage of ANSA in different years. There is not a clear pattern in the evolution of the corruption index, with several peaks from time to time. For example, the peak in May 2019 (circled in red) corresponds to a massive scandal that brought the mayor of Legnano (Milan) to be arrested for abuse of office.

[8] I consider false positive the articles that talk about corruption but do not involve a local politician directly.

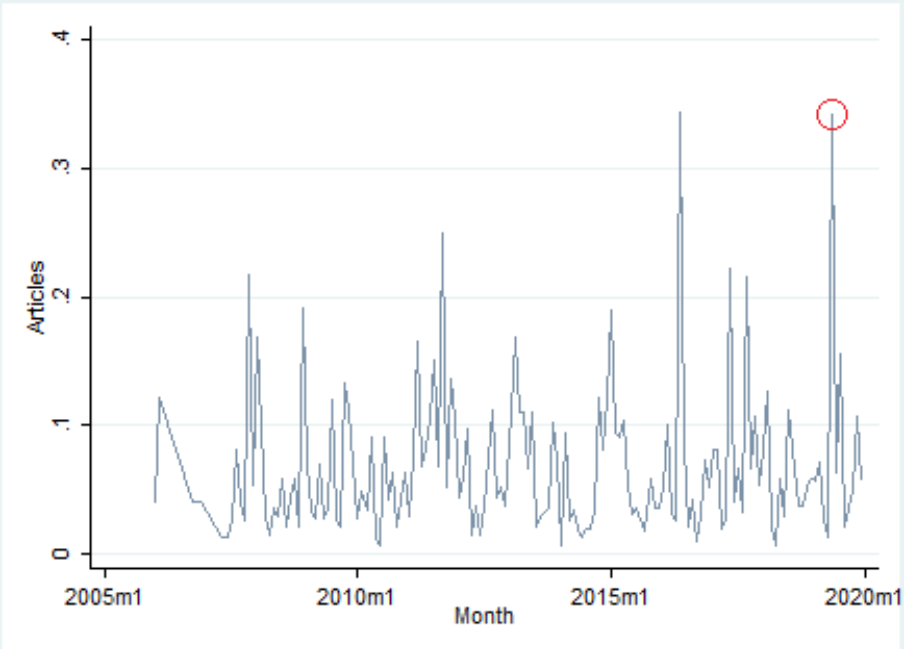
Figure 2 Descriptive analysis of scandals



(a) Distribution of the articles (b) Geographic area

Notes: The figure reports a descriptive analysis of the sample of corruption articles. The left figure shows the share of articles that are related to politicians in office in provincial capitals, mayors, and male politicians. The right figure shows the geographic distribution of the articles of corruption in comparison with the sample of cities in the analysis.

Figure 3 Distribution of articles at the monthly frequency



Notes: The figure reports the evolution of a corruption index at the monthly frequency. The index is built as follows: $corr_{t,y} = \frac{n_t^C}{N_y} \times 1000$ where n_t^C, where n_t^C is the number of corruption articles in month t of a given year, while N_y is the total number of ANSA articles published by Factiva in that particular year. The red circle is an example of a scandal, which occurred in Legnano (Milan) in May 2019.

3.3 Turnout To analyze the correlation between local scandals of corruption and turnout, I collected data about the turnout at the municipal level in the period 2006-2020 for all the three main types of elections: municipal, regional and national elections.[9] Turnout is measured by the ratio between valid votes and eligible citizens in a particular municipality.

Table 1 reports summary statistics of turnout. On average, citizens participate more in the national elections, consistent with other findings in the literature (Reif and Schmitt, 1980), (Marsh, 1998). Instead, among the local elections, turnout is generally higher at the municipal elections rather than at the regional. Columns 3-5 shows that there is a consistent heterogeneity in political participation within different geographic areas. Overall, turnout in the north and the center of Italy is similar in every type of election. Instead, in the south of Italy turnout is systematically lower for what concerns the national and regional elections, while it is slightly higher than the national average in the municipal elections.

Table 1 Descriptive analysis: Turnout

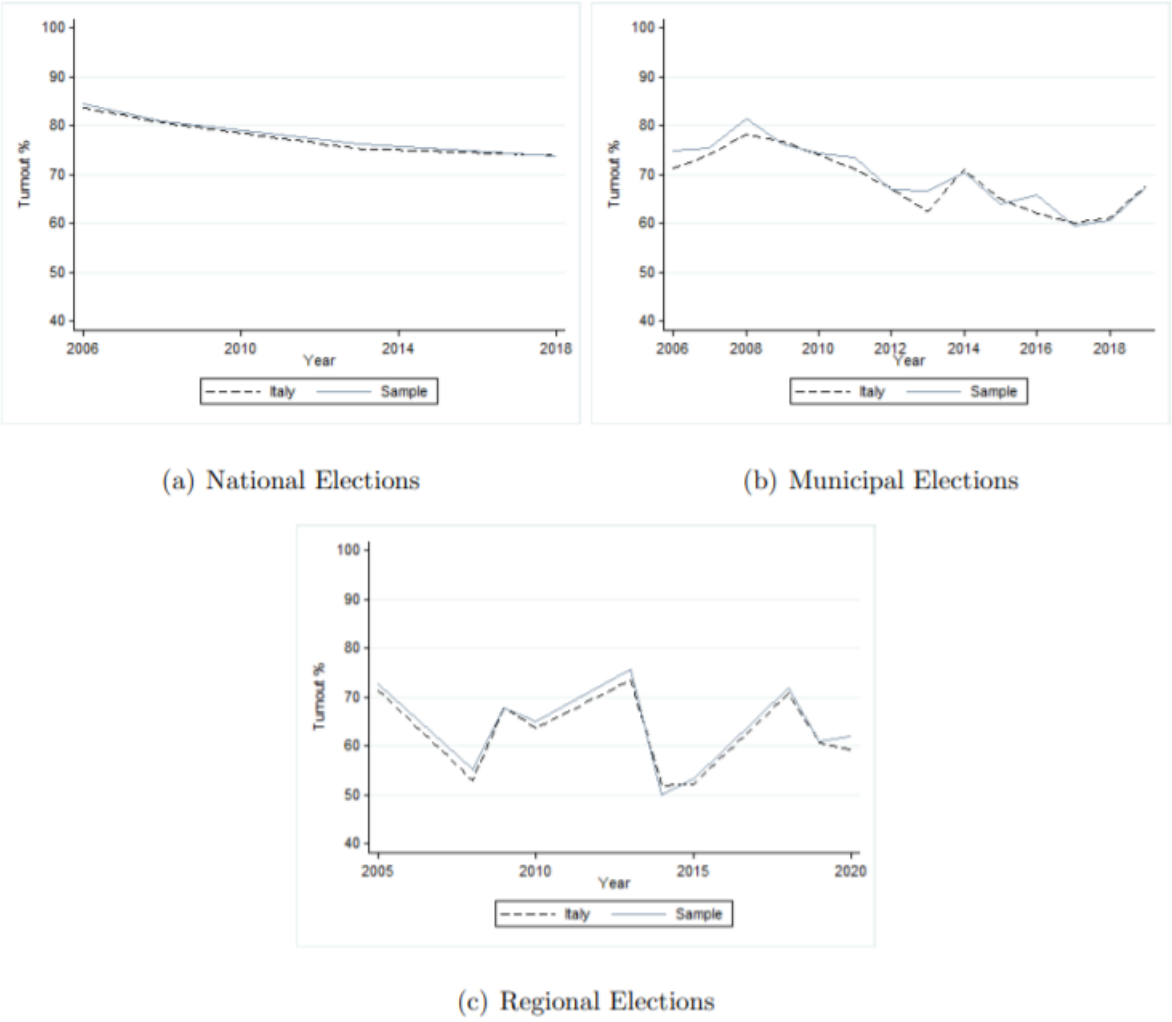
Type of Election	Mean	North	Center	South	Standard Deviation	Minimum	Maximum	Number of Obs
National	0.787	0.818	0.809	0.736	0.067	0.526	0.937	1,480
Municipal	0.704	0.677	0.706	0.734	0.079	0.482	0.885	1,136
Regional	0.657	0.680	0.655	0.626	0.105	0.250	0.863	1,274

Notes: The table shows the descriptive statistics of turnout for the sample of municipalities for the period in analysis, 2006–2020. Summary statistics are reported for the three main types of elections: national, regional, and municipal elections.

Figure 4 reports the evolution of turnout over time. To show the representativeness of the sample, I make a comparison between the turnout in the sample of cities (*grey line*) and the turnout in the pool of all Italian municipalities (*black dashed line*). The two lines overlap nicely in all three graphs, showing that turnout in the sample is indeed representative of the overall evolution of turnout in Italy. Turnout at the national elections is smoothly decreasing over time. Turnout at municipal elections has decreased over time as well, even if the pattern is slightly more irregular. For what concerns regional elections, there is not a clear pattern. This is because every year only a few regions vote. Therefore, the high irregularity captures some geographic heterogeneity within different regions.

[9] All electoral data are provided by the Italian Ministry of Internal Affairs. Turnout at municipal elections is built using valid votes at the first ballot; turnout at national elections is build using valid votes in the elections of *Camera*, one of the two branches of the Italian Parliament.

Figure 4 Evolution of turnout



(a) National Elections

(b) Municipal Elections

(c) Regional Elections

Notes: The figure shows the evolution of turnout, in the three types of elections, over the period of the analysis, 2006-2020. The grey line represents the average turnout in the sample of cities, while the dashed black line corresponds to the average turnout in all Italian municipalities.

3.4 Sociodemographic and economic variables To control for other factors influencing voter turnout, I introduce additional explanatory variables that have become standard in the turnout literature (Escaleras et al., 2012). Table 2 reports the summary statistics of the set of covariates that I use in the empirical analysis. The set of explanatory variables consists of a panel that covers the period 2006-2020. All the information is collected at the municipal level, except for *Unemployment*, that is, the unemployment rate at the province level, and *College Degree*, that is the share of the population with at least a college degree at the regional level. The variable *Population>18*, which represents the share of the population in voting age, is included to test the Downsian argument that a larger voting population reduces the probability that one vote will be decisive, in turn reducing turnout (Downs et al., 1957).

The other variables included are demographic and socioeconomic characteristics by municipality: total population in the city, the proportion of the population sixty-five years old and over, the share of women resident in the city, and the average income in the city. All the variables are taken from the Italian National Institute of Statistics (ISTAT), except for *Avg Income*, which was obtained from the Italian Ministry of Economics and Finance.[10] Figures 7-8 of Appendix I provide evidence of the balanceness of the sample of cities concerning the sociodemographic controls.

Table 2 Summary Statistics of covariates

	Mean	Standard Deviation	Minimum	Maximum	Number of Obs
Population	46,004	33,256	14,360	258,536	3,849
Share of Women	0.517	0.008	0.482	0.544	3,849
Population> 18	0.813	0.047	0.642	1.035	3,849
Population> 65	0.206	0.042	0.079	0.303	3,849
Avg Income	22,603	3,482	13,175	40,000	3,680
Unemployment	0.100	0.054	0.019	0.295	3,857
College Degree	0.148	0.027	0.094	0.233	3,890

Notes: The table shows the descriptive statistics for the sample of municipalities, for the period in analysis, 2006–2020. *Share of Women*, *Population>18*, and *Population>65* are measured as shares out of the municipal population. *Unemployment* is the unemployment rate at provincial level. *College Degree* is the share of college laureate at regional level.

4 Empirical Design

To analyze the correlation between local scandals of corruption and turnout at the municipal level, In the baseline specification, I propose to estimate a city fixed-effects model. At first, I aggregate together the electoral data from all the types of elections. There are multiple reasons behind this choice. First, aggregating the electoral data leads to an increase in the sample size, allowing for a larger variability in the corruption data and possibly for more precise estimates. Then, it is not immediate that local corruption should have a different effect in local elections and national elections. While I have found nothing in the literature that fosters this idea, such a hypothesis is still plausible. In order to test the presence of heterogeneity in the voting behavior in different type of elections, I will separately estimate the model for local elections and national elections in section 5.2.

[10]The variable is a proxy of average income built using data about aggregate taxable income at the annual frequency by municipality

The baseline specification is given by the following fixed effects model:

$$\text{Turnout}_{i,t,j} = \beta_0 + \beta_1 \text{Scandal}_{i,t,t-1} + \beta_2 \text{X}_{i,t} + \beta_3 \text{Contemporary}_{i,t,j} + \delta_t + \theta_j + \gamma_i + \varepsilon_{i,t} \quad (1)$$

The dependent variable is TURNOUT in city i, in year t, in type of election j. The variable of interest, SCANDAL, is a dummy that takes value one if there was a corruption scandal in city i, either in the year of the elections (until the election day) or in the year before. Therefore, I am estimating the correlation between turnout and corruption in the short term by considering scandals relatively close in time to the elections.

The $\mathbf{X}_{i,t}$ vector includes all the socio-economic and demographic variables presented in section 3.4. I also include type of election fixed-effects θ_j to account for the systematic difference in turnout observed in different types of elections. Furthermore, in Italy, it is common to vote on the same day for more than one election. At contemporary elections citizens tend to be more attracted to the poll, with a positive effect on turnout (Escaleras et al., 2012),(Reif and Schmitt, 1980). I control for this potential source of bias by including a dummy variable that assigns value one to contemporary elections in the model. Finally, I control for year fixed effects δ_t and city fixed effects γ_i.

In addition to equation 1, I propose a second specification, in which I include fixed effects by geographic area (North, South, Center of Italy) in the place of the city fixed-effects. While the city FE model is the benchmark for the analysis, it might suffer from some data limitations. That is, the independent variable SCANDAL might vary too little within the same unit over time. When the independent variable changes little overtime for the same unit, estimating a model with individual fixed effects could be problematic. The city FE model could encounter a similar problem since many municipalities in the sample are never treated with a corruption scandal, and most of the treated cities only faced few scandals. If such a problem exists, which I cannot exclude, the model with geographic area fixed-effects may be more convenient, even though it does not consider time-invariant characteristics at the municipal level that might be crucial determinants of turnout.

Finally, in section 5.3 I test the hypothesis that corruption and turnout correlate differently in provincial capitals and not provincial capitals. To detect such heterogeneity, I include in equation 1 an interaction term PROVINCE×SCANDAL.

5 Results

5.1 Baseline model The empirical results from the estimation of equation 1 are presented in table 3.[11] To provide insight into the source of variation in the model, I start with the baseline model estimated using ordinary least squares (OLS), with no fixed effects in Column 1. Column 2 presents the results of a year fixed effects and type of election fixed effects model, where I also control for contemporary elections. The coefficient of interest is negative and statistically significant, showing a negative correlation between scandals and turnout. Column 3 presents the results of a city-fixed effects model, where I also include the vector of socioeconomic and demographic controls. When estimating the model with city fixed effects, the coefficient of interest is negative yet not significant at the standard levels. This result suggests that most of the variation in the data is due to unobservable and time-invariant characteristics at the municipal level that determine both corruption and turnout. Accordingly, the negative correlation evidenced in Columns 1-2 could be entirely due to cities that are both systematically more corrupted and where political involvement is low. Overall, after controlling for fixed effects at the municipal level, there is no significant correlation between political corruption and turnout.

Column 4 reports the results of estimating a zone FE model, where I aggregate the cities according to the geographical area of belonging (North, South, or Center of Italy). In this specification, the coefficient of interest is negative and statistically significant at the 5 percent level. On average, after controlling for sociodemographic variables, in municipalities treated with a corruption scandal in the year preceding the elections, turnout is 0.87 percent lower. Theoretically, this result would be consistent with the *retrospective voting model* (Persson and Tabellini, 2002), where corruption works as a signal for the entire political class, leading voters to negatively update their beliefs and to reduce their participation in equilibrium. Lastly, Column 5 reports the results of a placebo test, where I regress turnout over the scandals that happen the year after the elections, including zone FE.[12] The estimated coefficient is not statistically significant, providing some robustness for the results found in Column 4.

[11] In all the regressions the variables *Population* and *Income* enter in log

[12] The placebo test is built in the following way. First, I randomly generate some fictitious scandals in the cities that were never treated. Then, for what concerns municipalities with at least one scandal, I consider in the regression the scandals that happen in the year immediately after the election.

As discussed in section 4, the city FE model is the benchmark of the analysis. However, estimating it could be problematic due to some limitations in the structure of the data. Even if the results of Columns 4-5 point in different directions, a given result is that I can exclude the existence of a positive correlation between corruption and turnout. Given the considerable disagreement present in the literature, this is a relevant claim that contributes to disproving the idea that political corruption increases turnout, either because corruption increases the expected returns for being in office (Lacombe et al., 2016), (Escaleras et al., 2012), or because citizens mobilize on to support clean governments (Bratton et al., 2005).

Turning to the socioeconomic control variables, I find results in line with the literature. The size of the share of the population in voting age (Share$>$ 18) exerts a negative and significant influence on voter turnout in Column 4, supporting the Downsian hypothesis (Downs et al., 1957), (Escaleras et al., 2012): the smaller the probability of an individual's vote being decisive, the less incentive there is to participate. On the other hand, in the city FE model, I find a positive coefficient. Once I account for unobservables characteristics at the city level, a larger share of citizens in voting age increases turnout. On average, an increase in population by 1 percent is associated with a reduction in turnout of 2.07 percent, consistent with (Van Houwelingen, 2017), while a 1 percent increase in average income leads to a 3.32 percent increase in turnout. Furthermore, I find a negative and significant correlation between unemployment and turnout. Here it should be noted that the data for the unemployment rate are collected at the province level. The interpretation of the result is that cities located in provinces with higher unemployment turn out less. Lastly, I find a positive and significant coefficient associated with contemporary elections, revealing that citizens are indeed more attracted to the poll when there are multiple elections on the same day.

Table 3 Empirical Results: turnout in every type of election

	(1) OLS	(2) Fixed Effect	(3) City FE	(4) Zone FE	(5) Placebo
Scandal	-0.0445***	-0.0145***	-0.00323	-0.00881**	-0.00171
	(0.00649)	(0.00408)	(0.00440)	(0.00430)	(0.00303)
Population			-0.207***	-0.0133***	-0.0141***
			(0.0438)	(0.00251)	(0.00247)
Income			0.332***	0.0412***	0.0403***
			(0.0621)	(0.0125)	(0.0125)
Share of women			-0.660	0.247	0.243
			(0.529)	(0.220)	(0.220)
Share>65			-0.0899	-0.0568	-0.0591
			(0.138)	(0.0592)	(0.0590)
Share>18			0.151***	-0.115***	-0.115***
			(0.0497)	(0.0354)	(0.0354)
Contemporary		0.0799***	0.0881***	0.0782***	0.0783***
		(0.00294)	(0.00308)	(0.00320)	(0.00320)
Unemployment			-0.143**	-0.176***	-0.177***
			(0.0604)	(0.0486)	(0.0485)
Observations	3535	3535	3300	3300	3300
Year	No	Yes	Yes	Yes	Yes
Election	No	Yes	Yes	Yes	Yes
City	No	No	Yes	No	No
Zone	No	No	No	Yes	Yes

Notes: Model I, reports OLS regression of turnout over scandal. Model II, control for year fixed-effects type of election fixed-effects, and contemporary elections. Model III, estimate equation 1. Model IV, includes geographic area fixed-effects. Model V, is a placebo test that build again on model IV.
Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.2 Heterogeneity between local and national elections In this section, I explore the hypothesis that the impact of local corruption on turnout differs according to the type of election in the analysis. While it is pretty intuitive that local scandals might matter for turnout in the local elections, it is not clear whether a correlation could also be expected in the national elections. Two main hypotheses are brought into the discussion. On the one hand, a local scandal could influence turnout only at the local elections, either positively or negatively, according to the different channels proposed in the literature. On the other hand, local scandals could influence turnout in all elections. In this scenario, by changing the voters'

behavior systematically, local corruption would have some *spillover effect* for politics at the national level.

To shed light on the discussion, I divide the sample into local elections and national elections. Then, I estimate equation 1 for both types of elections. In the *local elections*, together with the municipal elections, I also include the regional elections for two important reasons. First, many local politicians at the city level also have a political position in regional governments.[13] Second, it is reasonable to think that there could be spillover effects within regions. That is, a region with many scandals of corruption could be considered systematically more corrupted by the citizens. This, in turn, might impact turnout in the regional elections.

Table 4 Empirical Results: Different types of elections

	(1) City FE Loc	(2) City FE Nat	(3) Zone FE Loc	(4) Zone FE Nat	(5) Placebo Loc	(6) Placebo Nat
Scandal	-0.00684	0.00179	-0.0103*	-0.00748*	-0.00271	-0.000669
	(0.00611)	(0.00276)	(0.00578)	(0.00392)	(0.00407)	(0.00250)
Population	-0.134*	-0.107***	-0.0204***	-0.00875***	-0.0213***	-0.00941***
	(0.0757)	(0.0261)	(0.00348)	(0.00200)	(0.00341)	(0.00196)
Avg Income	-0.146	-0.0669*	0.0441**	0.0597***	0.0436**	0.0583***
	(0.103)	(0.0372)	(0.0178)	(0.0101)	(0.0178)	(0.0101)
Share of Women	-1.244*	1.329***	0.513*	-0.0327	0.498*	-0.0287
	(0.751)	(0.429)	(0.288)	(0.217)	(0.289)	(0.216)
Share>65	0.248	-0.223***	-0.0789	-0.0305	-0.0804	-0.0331
	(0.289)	(0.0683)	(0.0765)	(0.0659)	(0.0763)	(0.0656)
Share>18	0.150	-0.0452	-0.344***	-0.0178	-0.343***	-0.0192
	(0.140)	(0.0535)	(0.0430)	(0.0399)	(0.0429)	(0.0400)
Unemployment	0.00427	-0.0890**	0.0831	-0.475***	0.0802	-0.475***
	(0.0930)	(0.0401)	(0.0674)	(0.0444)	(0.0671)	(0.0444)
Contemporary	0.135***	0.0140***	0.122***	0.000286	0.122***	0.000419
	(0.00510)	(0.00321)	(0.00419)	(0.00338)	(0.00419)	(0.00337)
Observations	1838	1462	1838	1462	1838	1462
Year	Yes	Yes	Yes	Yes	Yes	Yes
Election	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	No	No	No	No
Zone	No	No	Yes	Yes	Yes	Yes

Notes: All the specifications include the set of socioeconomic and demographic controls as well as year fixed-effects and type of election fixed effects. Model I, controls for city fixed-effects and consider only turnout at the regional and the municipal elections. Model II, controls for city fixed-effects and consider only turnout at the national elections. Model III, controls for geographic areas fixed-effects and consider only turnout at the regional and the municipal elections. Model IV, controls for geographic areas fixed-effects and consider only turnout at the national elections. Model V, is a placebo test that build on Model III. Model VI, is a placebo test that build on Model IV.
Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[13] This is fairly common for smaller cities, typically not Province capitals.

The empirical results are reported in table 4. Column 1-2 show the results of estimating equation 1, for both local and national elections. The correlation is negative yet not statistically significant in local elections, while it is slightly positive and statistically insignificant in national elections. The results of the city FE model show that there is no significant heterogeneity in the way local corruption relates to turnout at the municipal level. Columns 3-4, instead, show the results of the zone FE model. In these specifications, the coefficient of interest is both negative and statistically significant for all elections. The negative correlation seems to be stronger for local elections. On average, in municipalities treated with a local scandal, turnout drops by 1.03 percent if the next election is local, while only by 0.75 percent if the next election is national. The zone FE results confirm the absence of heterogeneity in how local corruption relates to turnout in different types of elections. Lastly, Columns 5-6 report the results of placebo tests to provide robustness for the results of Columns 3-4.[14]

5.3 Heterogeneity in size and administrative power At least in principle, there could be another source of heterogeneity: corruption and turnout might correlate differently according to the city's size and administrative power. Previous works in the literature have focused on the relationship between sociopolitical characteristics and turnout. For instance, Van Houwelingen (2017) shows that turnout tends to be lower in bigger cities. Additionally, Dahl and Tufte (1973) and Monroe (1977) study the variation of turnout according to the administrative importance of a city and the degree of urbanization. Following this strand of literature, also the relation betweeen corruption and turnout might differ according to the similar characteristics. To test for this hypothesis, I exploit the variation related to the status of provincial capitals, which is generally assigned to the most important and biggest cities in Italy. I estimate again equation 1, including the interaction term SCANDAL $\times$ PROVINCE, where PROVINCE is a dummy that takes value 1 if the city is a provincial capital. The empirical results are reported in table 5.

[14]The placebo is estimated in the same way as in section 5.1

Table 5 Empirical Results: provincial capitals vs not provincial capitals

	(1) OLS	(2) City FE	(3) Zone FE	(4) Not Prov	(5) Prov	(6) Placebo
Scandal	-0.0459***	-0.00736	-0.0168***	-0.0159***	-0.00238	-0.00351
	(0.00939)	(0.00618)	(0.00604)	(0.00611)	(0.00604)	(0.00355)
Interaction	0.00273	0.00979	0.0167**			
	(0.0125)	(0.00863)	(0.00812)			
Population		-0.195***	-0.0151***	-0.0268***	0.00493	-0.0275***
		(0.0433)	(0.00249)	(0.00386)	(0.00449)	(0.00386)
Income		0.342***	0.0537***	0.0509***	0.0883**	0.0502***
		(0.0617)	(0.0122)	(0.0139)	(0.0367)	(0.0140)
Share of women		-0.549	0.0697	0.245	-0.625	0.249
		(0.529)	(0.215)	(0.238)	(0.616)	(0.238)
Share>65		-0.0712	0.00727	-0.0266	0.121	-0.0258
		(0.136)	(0.0553)	(0.0592)	(0.151)	(0.0591)
Share>18		0.145***	-0.145***	-0.148***	-0.0477	-0.150***
		(0.0503)	(0.0346)	(0.0376)	(0.0933)	(0.0376)
Contemporary		0.0882***	0.0782***	0.0806***	0.0753***	0.0807***
		(0.00305)	(0.00320)	(0.00354)	(0.00768)	(0.00354)
Observations	3535	3325	3325	2634	691	2634
Year	No	Yes	Yes	Yes	Yes	Yes
Election	No	Yes	Yes	Yes	Yes	Yes
City	No	Yes	No	No	No	No
Zone	No	No	Yes	Yes	Yes	Yes

Notes: Model I, reports an OLS regression of turnout over scandal, where I include the interaction term *Scandal*×*Province*. Model II, control for year fixed-effects, type of election fixed-effects, and city fixed-effects. Model III, control for geographic areas fixed-effects, in the place of city fixed-effects. Model IV and V estimate again Model III, keeping only the sample of not provincial capitals and provincial capitals (in order). Model VI, is a placebo test that build again on model IV.
Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Column 1 reports the results of the baseline model estimated with ordinary least squares. Column 2 shows the city FE model results. Consistent with the results in section 5.1 and 5.2, both the interaction term and the variable of interest don't exert a significant effect on turnout. There is no heterogeneity in how scandals relate with turnout in provinces and not provinces after controlling for time-invariant characteristics at the municipal level.

Column 3 reports the results of estimating a zone FE model. The coefficient associated with the interaction term is positive and statistically significant. On average, an additional scandal

reported in a city that is not provincial capital leads to a reduction in turnout of 1.68 percent. The effect of an additional scandal in provincial capitals (given by the sum of the coefficients associated with SCANDAL and INTERACTION) might not be statistically significant overall. To test this hypothesis, I divide the sample into provinces and not provinces. I then estimate the zone FE model again. On average, an additional scandal in cities that are not provincial capitals is associated with a 1.59 percent decrease in turnout, while the effect is both statistically insignificant and very close to zero for what concerns provincial capitals. Lastly, Column 5 reports the results of a placebo test, where I kept the sample of non-provinces, and I control for geographic areas fixed-effects.

Overall, the results are not conclusive. The zone FE model results provide weak evidence of heterogeneous response to corruption scandals in cities between provinces and not provinces. However, after controlling for city FE, no statistically significant correlation appears.

6 Identification Issues and Data Limitations

The results reported in section 5 show mixed results. Overall, the city fixed-effects model shows no significant correlation between scandals and turnout. Instead, the zone fixed-effects model evidence a negative and significant correlation. Even if the former model is the benchmark, I cannot exclude that the estimations might suffer from some limitations of the corruption data (see section 4). The data collection strategy that I use is designed to limit the influence of false positives. However, it may disregard many false negatives. That is, there might be scandals of local corruption not included in the sample for several reasons. First, I cannot record scandals that occurred after a politician leaves office, which may indeed affect voters' behavior. Second, ANSA could fail in reporting some scandals, especially concerning small cities. These data limitations mechanically reduce the variability in the dataset, the reason why it might be problematic to rely on the city fixed-effects model estimations. A possible way to overcome the issue, at least in part, would be to extend the data collection to local newspapers. This would allow for better coverage of every city in the sample, enhancing the variability in the data.

On the other hand, a considerable limitation of the zone FE model is that it fails in controlling for time-invariant and unobservable characteristics at the municipal level which might be determinants for turnout. Furthermore, even if one relies on such estimates, it would be incorrect to

interpret the negative correlation as a causal effect. A potential problem of endongeneity could be related to the existence of reverse causality. That is, the lower the political participation in the city, the higher the incentive of local politicians to misbehave. An argument could be that a lower political involvement of the citizens is related to a minor control on the incumbent's activity, which in turn fosters corruption by decreasing the political accountability of the municipal government (Lederman et al., 2001).

A final concern comes from the measure of corruption that I use in the analysis. It is essential to mention that what I am measuring is the awareness of the phenomenon of corruption as it is reported by the media, which may differ from the actual amount of corruption in a certain city. Nonetheless, this measure of exposure is what matters if we want to study the impact on political outcomes, as it is the corruption perceived by the people (Giommoni, 2021).

7 Conclusions

In this paper, I study the effect of corruption exposure on electoral turnout, using data for Italian elections in the period 2006-2020. For the analysis, I focus on local scandals involving politicians in office in the municipal governments. By screening newspaper articles released by the leading Italian press agency, ANSA, I build a measure of local corruption for a sample of 370 Italian municipalities covering the entire period of analysis.

The main results evidence the absence of a statistically significant correlation between local corruption and turnout. This is consistent with previous findings in the literature (Stockemer, 2013), (Persson et al., 2013). Still, the results might not be conclusive, as the identification strategy could suffer from some data limitations. I propose a second empirical design that evidences a negative and significant correlation between local scandals and turnout rates. This result, consistent with a different view that has been proposed in the literature (Costas-Pérez, 2013), (Chong et al., 2015), might be interpreted in the light of the *retrospective voter model* (Persson and Tabellini, 2002), where corruption works as a signal for the entire political class, leading voters to update their beliefs negatively and to reduce their participation in equilibrium. Moreover, if I rely on the second specification, I cannot exclude that the negative correlation might be driven by reverse causality. Even if the previous results go in different directions, they seem to disprove the idea that corruption exposure positively affects turnout, a hypothesis

supported by a consistent number of scholars in the previous literature (Escaleras et al., 2012), (Lacombe et al., 2016), (Bratton et al., 2005).

I then test two plausible sources of heterogeneity. First, I test the hypothesis that exposure to local corruption might impact turnout differently according to the type of election in the analysis. No heterogeneous effect between local and national elections is found in the data. Second, I test whether corruption exposure and turnout correlate differently in the provinces and not provinces. While the baseline specification provides evidence against the hypothesis, the second specification evidences a negative and significant correlation in cities that are not provincial capitals, while a negligible effect concerning provincial capital. This result suggests that such an heterogeneous effect might indeed be present. Future studies should provide new evidence to shed light on the discussion.

A final consideration is about the external validity of the results. These results concern the Italian context. The extent to which these may be general depends on an important remark. Italy is an essential outlier among OECD countries in terms of its corruption level. Therefore, the effects of corruption may be different elsewhere, for instance, because voters have different prior beliefs about corruption levels or because those interested in actively participating in politics have other incentives.

References

Bianca Clausen, Aart Kraay, and Zsolt Nyiri. Corruption and confidence in public institutions: Evidence from a global survey. *The World Bank Economic Review*, 25(2):212–249, 2011.

James Caillier. Citizen trust, political corruption, and voting behavior: Connecting the dots. *Politics & Policy*, 38(5):1015–1035, 2010.

Claudio Ferraz and Frederico Finan. Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes. *The Quarterly journal of economics*, 123(2):703–745, 2008.

Elena Costas-Pérez, Albert Solé-Ollé, and Pilar Sorribas-Navarro. Corruption scandals, voter information, and accountability. *European Journal of Political Economy*, 28(4):469–484, 2012.

Eric Avis, Claudio Ferraz, and Frederico Finan. Do government audits reduce corruption? estimating the impacts of exposing corrupt politicians. *Journal of Political Economy*, 126 (5):1912–1964, 2018.

Aksel Sundström and Daniel Stockemer. Quality of government affect voter turnout in the european regions. *QoG Working Paper Series*, 2013(6):6, 2013.

Daniel Stockemer. Corruption and turnout in presidential elections: A macro-level quantitative analysis. *Politics & Policy*, 41(2):189–212, 2013.

Michael T Rock. Corruption and democracy. *The Journal of Development Studies*, 45(1):55–75, 2009.

J Heidenheimer Arnold, Michael Johnston, and Victor T LeVine. Political corruption. a handbook. *New Brunswick/London*, 1989.

John G Matsusaka. Explaining voter turnout patterns: An information theory. *Public choice*, 84(1-2):91–117, 1995.

Anthony Downs et al. An economic theory of democracy. 1957.

Alberto Chong, Ana L De La O, Dean Karlan, and Leonard Wantchekon. Does corruption information inspire the fight or quash the hope? a field experiment in mexico on voter turnout, choice, and party identification. *The Journal of Politics*, 77(1):55–71, 2015.

Elena Costas-Pérez. The effects of political corruption on turnout: Mobilization or disaffection. *Universitat de Barcelona & Institut d'Economia de Barcelona (IEB)*, 2013.

Stephen Ansolabehere, Shanto Iyengar, Adam Simon, and Nicholas Valentino. Does attack advertising demobilize the electorate? *American political science review*, pages 829–838, 1994.

Tatiana Kostadinova. Abstain or rebel: Corruption perceptions and voting in east european elections. *Politics & Policy*, 37(4):691–714, 2009.

Donald J Lacombe, R Morris Coats, William F Shughart II, and Gökhan Karahan. Corruption and voter turnout: a spatial econometric approach. *Journal of Regional Analysis & Policy*, 46(2):168–185, 2016.

Monica Escaleras, Peter T Calcagno, and William F Shughart. Corruption and voter participation: Evidence from the us states. *Public Finance Review*, 40(6):789–815, 2012.

Michael Bratton, Robert Mattes, and Emmanuel Gyimah-Boadi. *Public opinion, democracy, and market reform in Africa*. Cambridge University Press, 2005.

Mikael Persson, Maria Solevid, and Richard Öhrvall. Voter turnout and political equality: Testing the 'law of dispersion'in a swedish natural experiment. *Politics*, 33(3):172–184, 2013.

Transparency international. 2020.

Alberto Vannucci. The controversial legacy of 'mani pulite': A critical analysis of italian corruption and anti-corruption policies. *Bulletin of Italian politics*, 1(2):233–264, 2009.

Alberto Vannucci. L'evoluzione della corruzione in italia: evidenza empirica, fattori facilitanti, politiche di contrasto. *MERLONI, F.; VANDELLI, L.(a cura di), La corruzione amministrativa. Cause, prevenzione e rimedi. Firenze*, page 37, 2010.

Autorità Nazionale Anticorruzione. La corruzione in italia (2016-2019). numeri, luoghi e contropartite del malaffare. 2019.

Tommaso Giommoni. Exposure to corruption and political participation: Evidence from italian municipalities. *European Journal of Political Economy*, 68:102000, 2021.

Gianmarco Daniele, Sergio Galletta, and Benny Geys. Abandon ship? party brands and politicians' responses to a political scandal. *Journal of Public Economics*, 184:104172, 2020.

Marco Le Moglie and Gilberto Turati. Electoral cycle bias in the media coverage of corruption news. *Journal of Economic Behavior & Organization*, 163:140–157, 2019.

Lucia Rizzica and Marco Tonello. Exposure to media and corruption perceptions. *Bank of Italy Temi di Discussione (Working Paper) No*, 1043, 2015.

Daniel Lederman, Norman Loayza, and Reis Rodrigo Soares. *Accountability and corruption: political institutions matter*. The World Bank, 2001.

Pepijn Van Houwelingen. Political participation and municipal population size: A meta-study. *Local Government Studies*, 43(3):408–428, 2017.

Robert Alan Dahl and Edward R Tufte. *Size and democracy*, volume 2. Stanford University Press, 1973.

Alan D Monroe. Urbanism and voter turnout: A note on some unexpected findings. *American Journal of Political Science*, pages 71–78, 1977.

Autorità Nazionale Anticorruzione. Relazione annuale 2015. *Roma Senato della Repubblica*, page 340, 2016.

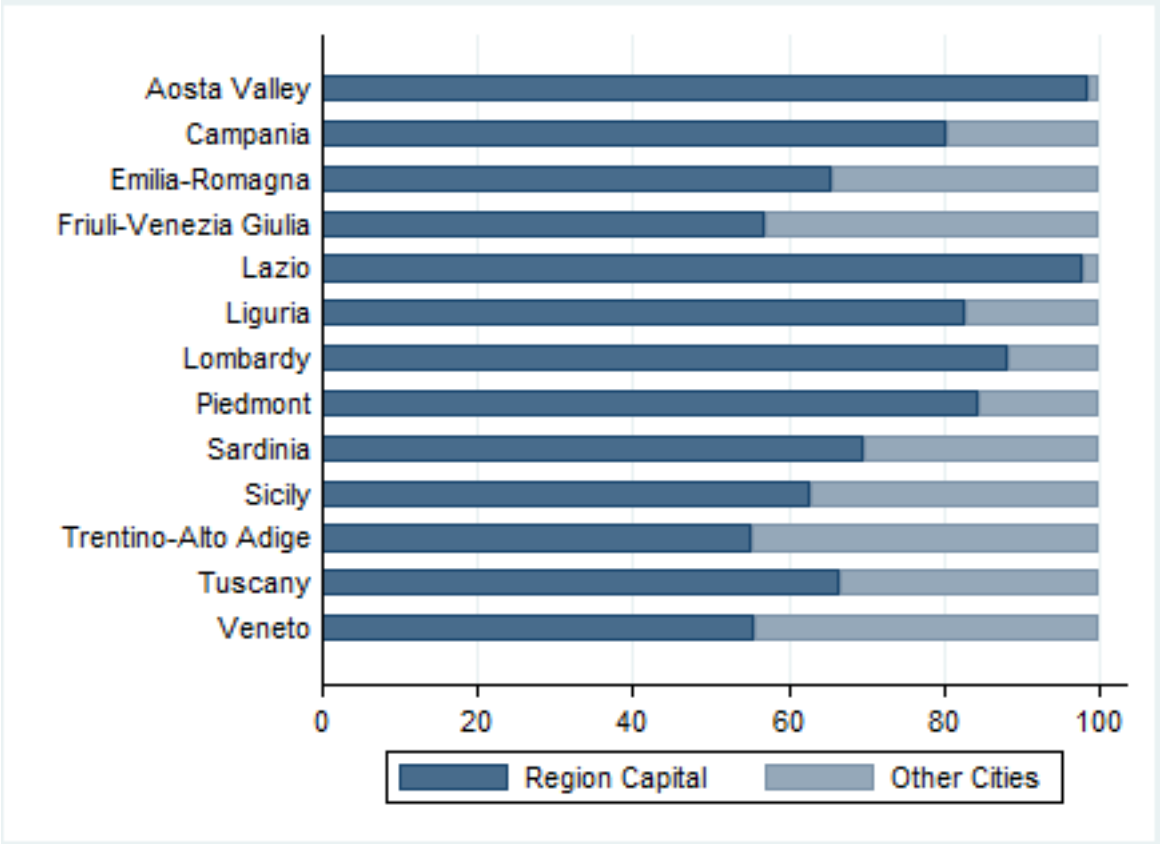
Karlheinz Reif and Hermann Schmitt. Nine second-order national elections–a conceptual framework for the analysis of european election results. *European journal of political research*, 8 (1):3–44, 1980.

Michael Marsh. Testing the second-order election model after four european elections. *British journal of political science*, pages 591–607, 1998.

Torsten Persson and Guido Tabellini. *Political economics: explaining economic policy*. MIT press, 2002.

Appendix I. Additional material

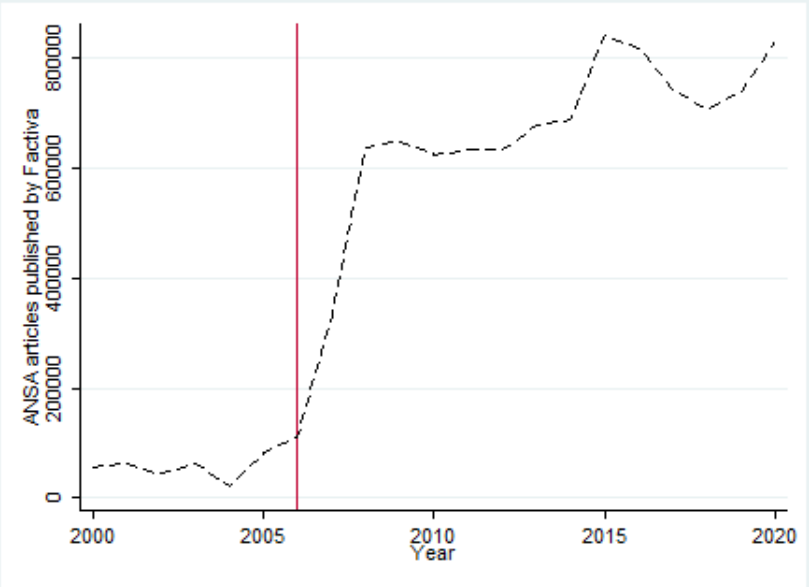
Figure 5 Share of ANSA articles related to regional capitals (2015)



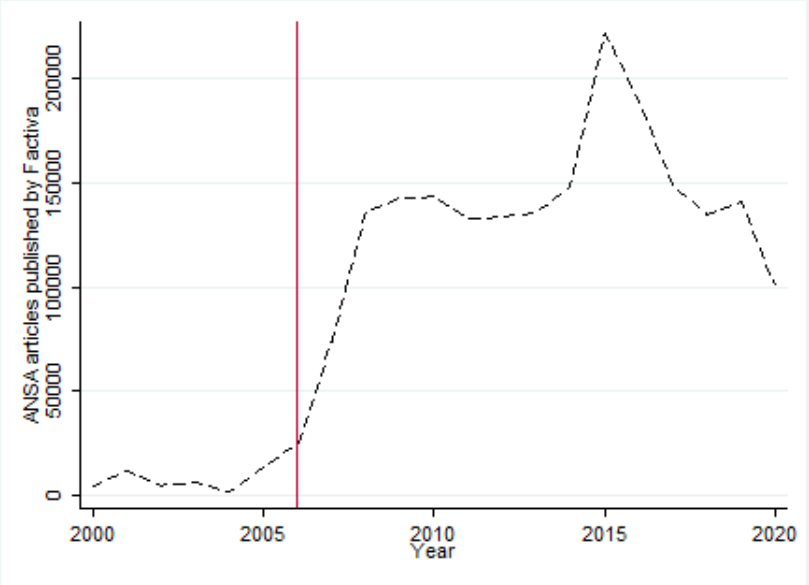
Notes: The figure provides evidence of the unbalanceness between the coverage of regional capitals and all the other cities. For each Italian region, I report the share of articles released by ANSA in 2015 related to the regional capital (SOURCE: Factiva).[a] In all the regions, the media coverage of the capital exceeds the threshold of fifty percent, and for most of them, it is even higher than sixty percent. In Aosta Valley and Lazio, the coverage of the region's capital is almost 100 percent.

[a]Factiva does not provide the geolocalization service for some Italian regions. These regions are not included in the graph.

Figure 6 Factiva's coverage in time



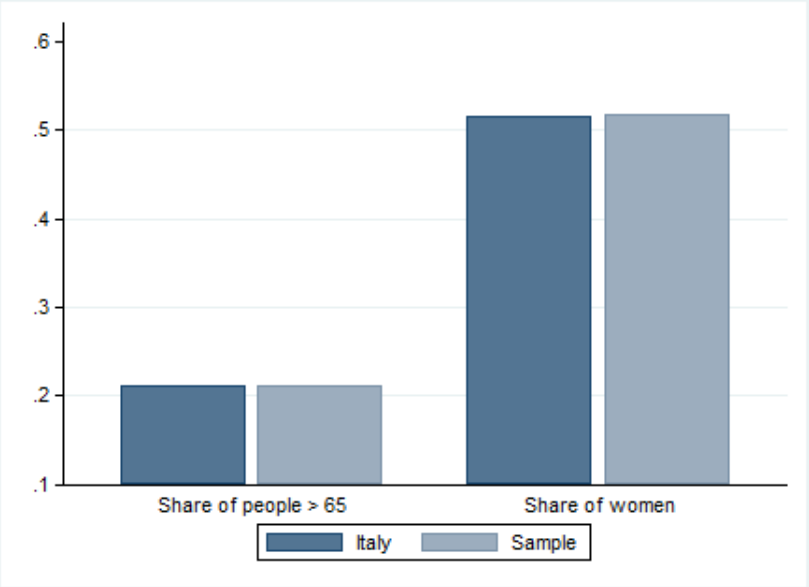
(a) Political Section



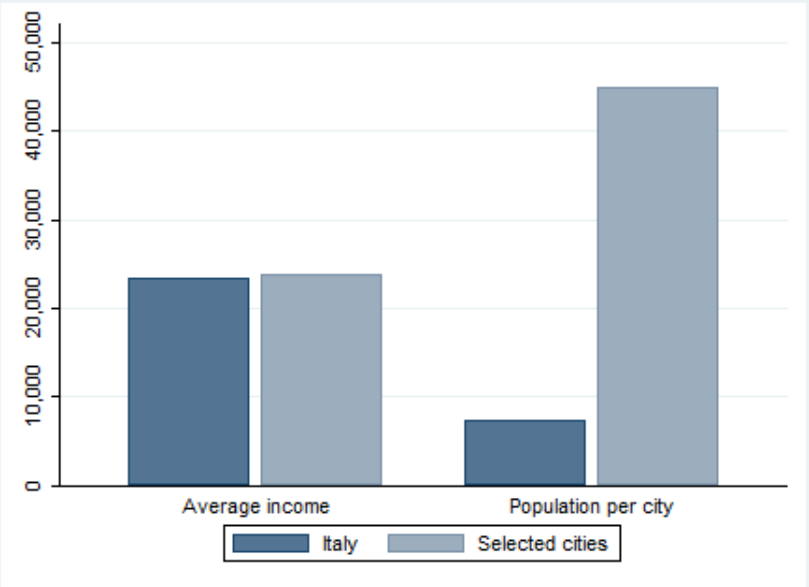
(b) Crime Section

Notes: The figure provides evidence of the evolution of Factiva's coverage over time. I report the number of articles released by ANSA that are included in the Factiva library for each year since 2000. Figure a) focus on articles in the *Political Section*, while figure b) on the *Political Crime Section*. The coverage improves significantly after 2006, which is marked with a vertical red line.

Figure 7 Comparison of covariates: sample of cities vs Italy



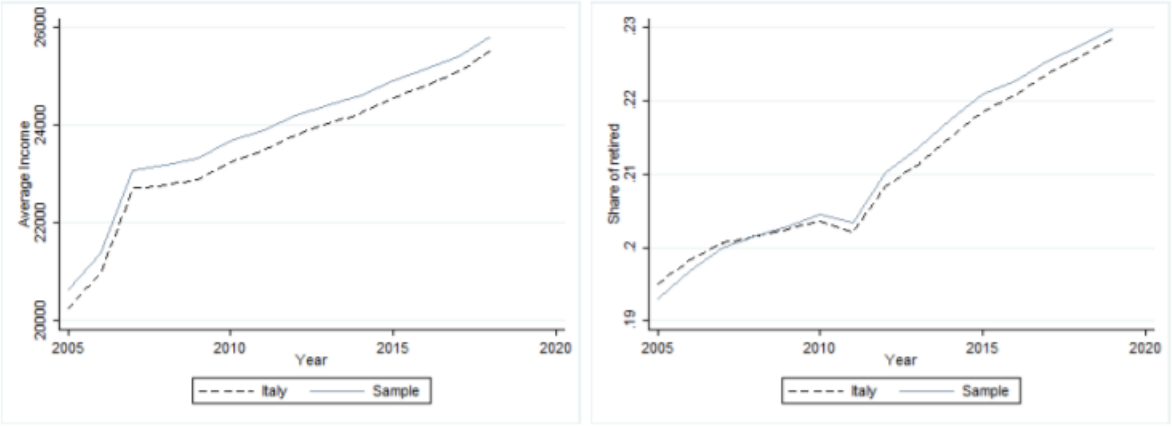
(a) Demographic variables



(b) Income and population

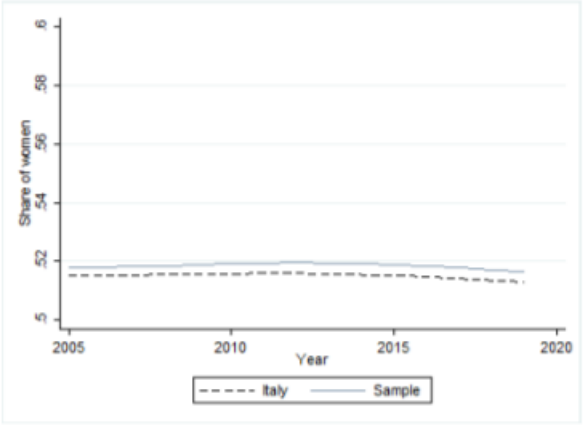
Notes: The figure shows the balanceness of the sample of cities concerning the set of sociodemographic and economic covariates. Figure a) reports the average share of women and the average share of retired people (proxied by the population older than 65) in the sample of cities and the pool of all Italian municipalities. Figure b) reports the average income and the average population. The averages are taken considering all the period 2006-2020. Overall, all the variables are well balanced, with the exception of population. This is by construction since I excluded all the municipalities with less than 20.000 inhabitants from the sample.

Figure 8 Evolution of sociodemographic covariates



(a) Average income

(b) Share of retired



(c) Share of women

Notes: The figure shows the balanceness of the sample of cities over time with respect to the average income, the share of women, and the share of retired (proxied by the population older than 65). The *grey line* is the average in the sample, while the *black dashed line* is the average within all Italian municipalities. Overall, the set of covariates is kept well balanced over time.

Table 6 Composition of the geographic areas

Zone	**Regions**	**N.Regions**
North	Emilia-Romagna, Liguria, Lombardy, Piedmont, Veneto	5
Center	Lazio, Marche, Tuscany, Umbria	4
South	Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardinia	7

Notes: The table describes how the Italian regions are divided into North, South, and Center of Italy. Four regions out of twenty are not present in the table, since they are excluded from the sample of cities due to the unavailability of electoral data.

Appendix II. Data collection

In this appendix, I discuss in more detail the procedure applied in the data collection. I adopt an automatic keyword-based procedure, structured in five steps:

1. **Dataset of politicians**. I created a dataset containing information of all local politicians (members of the municipal government) in office in the sample of cities in 2006-2020. For every politician, I know the name, surname, office title, the start date of the office, the end date of the office. This information comes from *Anagrafe Degli Amministratori Locali*, released every year by the Italian Ministry of Internal Affairs.

2. **Articles's downloading**. Through the portal Factiva, I downloaded articles released by ANSA and containing the name and surname of a local politician in the period when he/she was in charge.[15] Factiva uses an automatic classification system that classifies the articles according to their content. For the analysis, I focused on articles released in the *Political crime* section in the period 2006-2020.

3. **Geolocalization**. I geolocalized the selected articles based on places mentioned in the text. In particular, the text of the articles has a standard structure where the first word is usually the name of the place where the piece of news comes from. Figure 10 shows an example: the place (L'Aquila, capital of Abruzzo) is right at the beginning of the text. However, in most cases, the first word does not match exactly the city's name where the news comes from. Rather, in that position, ANSA usually reports the name of the relative province capital. In the phase of geolocalization, I only kept in the dataset the articles that report as a first word either the name of the municipality where the politician is in office or the name of the related province capital.

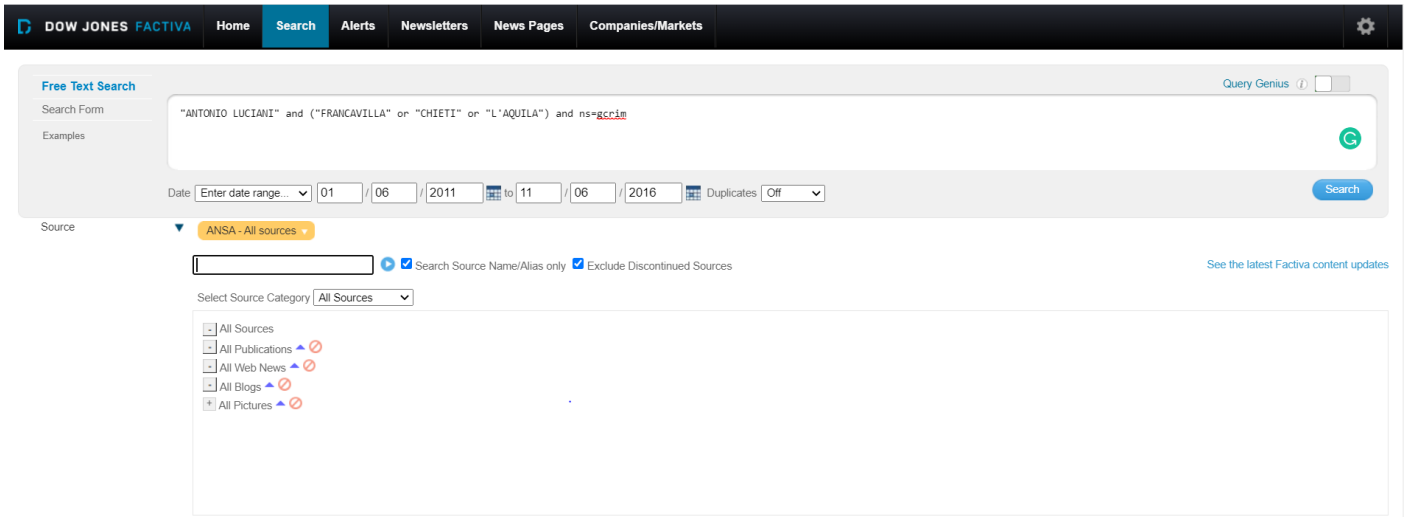
4. **Text Analysis** I use a *Bag of words* strategy in order to select the articles about political corruption. I identified three lists of keywords. First, a list of corruption keywords (e.g., the Italian translation of bribe, corruption, extortion, embezzlement, abuse of office). Then, a list of corruption-related verbs (arrested, investigated, accused, sent for trial, and synonyms). Finally, a list of political office titles (mayor, assessor, councilor, president of

[15] I use a python code to perform the extraction from the portal Factiva.

the city council). Using an automated code in python, I then selected the articles that contain in the same sentence the surname of the politician and/or his/her office title and at least one corruption-related keyword (either a noun or a verb).

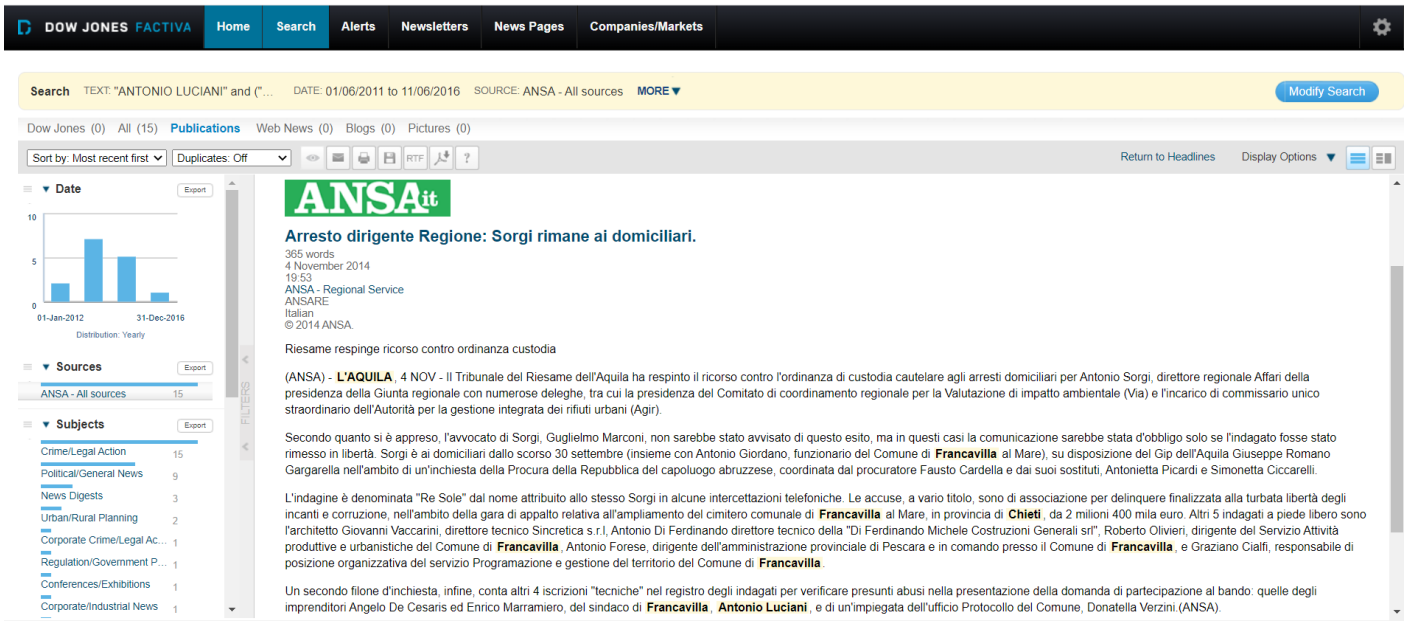
5. **Data cleaning**. In order to exclude the presence of false-positive, I manually checked the dataset of corruption articles that derive from the previous step.

Figure 9 Example of article search using Factiva



Notes: The figure reports an example of a search using the portal Factiva. I select all the available sources for ANSA. I search for articles released by ANSA in the period when the politician was in charge. I only look for articles that contain the name and surname of the local politician (in the example ANTONIO LUCIANI), and either the name of the city where he was in office (FRANCAVILLA) or the province where the city is located (CHIETI) or the region capital (L'AQUILA). The last requirement is a way to ease the procedure of geolocalization. The piece of code *ns=gcrim* indicates that I'm searching in the *Political Crime Section*.

Figure 10 Example of article extraction using Factiva



Notes: The figure reports an example of the structure of the articles released by ANSA. The first word is usually the place where the piece of news comes from (either the municipality, the related province capital, or the related region capital). ANSA always refers to people by using both the name and the surname. This is crucial for the data collection since I can search the full name of the politician (in the example ANTONIO LUCIANI) without losing a valuable piece of news. Other newspapers don't have this convenient feature, as they often refer to people only using their surname.