BAFFI RESEARCH CENTER

Working Paper Master Theses Series

Mapping Voters' Preferences with Digital Trace Data and Random Utility Models

Marvin Pappalettera | 2024

Università Bocconi
BAFFI
Centre on Economics, Finance and Regulation

Mapping Voters' Preferences with Digital Trace Data and Random Utility Models

Marvin Pappalettera

Supervisor: Massimo Morelli, Co-supervisor: Nicolò Cavalli

This research presents a novel methodological framework for applying the spatial theory of voting to digital trace data obtained from social media platforms. Traditional data collection methods often have limitations in capturing individuals' ideologies and political preferences, which are essential for the empirical application of this theory. On the other hand, recent advancements in online network ideological scaling techniques have allowed researchers to estimate the ideological positions of large samples of individuals based on their online activities. Nevertheless, there is still a significant gap in the literature when it comes to applying the spatial theory of voting to these data. This study aims to fill this gap by utilizing a novel dataset containing the ideological positions of hundreds of thousands of Twitter users. The study proposes a simultaneous model of party choice and abstention, wherein voters are positioned in a multi-dimensional ideological space and vote probabilistically as a function of their relative distance from the parties. The major contribution of this paper is an innovative estimation approach based on Maximum Likelihood Estimation. By treating the use of aggregate data as a measurement error problem, I demonstrate how to estimate this model without relying on individual-level choice data, making this framework ideal for working with digital traces. The analysis is divided into two parts. Firstly, the model is tested on the results of the 2022 national election in Italy. The results indicate that voters possess meaningful ideologies, and economic issues constitute the most relevant ideological dimension in explaining the election results. Secondly, by applying the spatial theory of voting in the framework of a "multi-class classification problem with aggregate data", the study demonstrates how to predict individual-level voting behavior accurately.

1 Introduction

Since the publication of *An Economic Theory of Democracy* by Anthony Downs in 1957, the spatial theory of voting has been dominant in explaining voting behavior. Based on a spatial conceptualization of politics, this theory states that it is possible to position political preferences in an abstract (ideological) space and that citizens vote for the party or policy alternative closest to them (Lipovetsky, 2022). The central idea is that a few latent dimensions, such as *left-right* and *liberal-conservative*, can capture variation in voters' preferences across many issues (e.g., abortion, immigration, taxation). These dimensions are often called *ideologies* and allow us to map individual preferences from a high-dimensional issue space onto a lower-dimensional one. As a result, it is possible to model voting behavior with just a few (usually one or two) ideological dimensions.

Traditionally, empirical applications of the spatial theory of voting have relied on surveys that ask respondents to place parties and themselves on ideological or issue scales.[1] The main drawback of this approach is that it can be problematic to uncover individuals' ideological stances and voting choices via conventional data collection methods. The topic's sensitive nature implies that issues commonly linked to survey methodology, such as response bias, are likely exacerbated. To address these limitations, this study proposes a novel methodological framework that utilizes digital trace data.

Nowadays, more and more people spend time on social media platforms like Twitter, Facebook, and Instagram. By interacting with them, they leave behind a "digital trace" of all their activities, including information such as the people they follow or interact with and their location. When this granular data is available for research, it offers valuable information on human behavior that more traditional data collection methods do not reveal. Platforms like Twitter are especially useful because they are often used to consume political content, making the digital behavioral traces of its users a source of information on public opinion on various topics of public discourse. Barberá (2015) was the first to show that it is possible to recover the liberal or conservative attitudes of millions of Twitter users in the US from their digital traces.

[1] An example is the following question from the "European Social Survey": "In politics people sometimes talk of "left" and "right". Where would you place yourself on this scale, where 0 means the left and 10 means the right?".

Unlike more traditional data collection methods, this data originates from an unobtrusive observation of individuals' behavior, making it appealing for two reasons. The first one is that it allows us to mine opinions on a massive scale. The second one is that, while some individuals may not feel comfortable sharing their opinions or ideological positions with an interviewer, they often unintentionally reveal this information as a by-product of their online activity.

Our study utilizes a novel dataset containing the ideological positions of hundreds of thousands of Twitter users obtained from Morales et al. (2022). To our knowledge, this is the first study to apply the spatial theory of voting to digital trace data. Firstly, we develop a model that simultaneously considers the choice among $J \geq 2$ parties and abstention. Voters are positioned in a bi-dimensional (ideological) space and vote probabilistically as a function of two components: (i) the relative weighted distance from each party, where the weights reflect the *salience* of each dimension, and (ii) a residual that captures the parties' *valence*.[2] Our analysis is then divided into two parts. In the first part, we estimate the salience weight of each dimension based on the 2022 national election in Italy. Our findings indicate that voters possess meaningful ideologies and that the dimensions we consider are significant in explaining the voting choices of the Italian electorate. Furthermore, this approach enables us to answer questions such as: "With what probability will citizens with ideal points at x vote for one candidate, the other, or abstain?" (McKelvey, 1975). Then, we demonstrate how this model can accurately predict individual voting behavior in a supervised learning framework.

The study is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the model in its general form. Section 4 describes the data. Sections 5 and 6 present the empirical results of the inferential and predictive applications, respectively. Section 7 concludes.

[2]The concept of *valence* is related to factors such as the party leader's charisma or the party's reputation that can affect voting choices.

2 Overview of spatial voting and other relevant literature

Hotelling (1929) and Smithies (1941) are traditionally accredited with the idea of spatial competition. They studied the optimum location of firms in a *linear* space. In this space, buyers of a single commodity are uniformly distributed along a single line of length l. The sellers, A and B, are free to move along this line, and everyone buys from the one closest to them. Under these assumptions, Hotelling (1929) found that there is a tendency for sellers to crowd together as closely as possible at the center of the distribution. If A were to settle at any point but the median, B would fix his location between A and the center, as close to A as possible, to maximize his profits. This incentive to "undercut" each other to capture as many buyers as possible drives the firms toward the center. Hotelling (1929) noted that the same tendency can be found in politics:

> The competition for votes between the Republican and Democratic parties does not lead to a clear drawing of issues, an adoption of two strongly contrasted positions between which the voter may choose. Instead, each party strives to make its platform as much like the other's as possible.

In *An Economic Theory of Democracy* (1957), A. Downs formalized this model in the context of competition between political parties, establishing spatial theory as a conceptual tool. He assumed that voters are distributed along a single *ideological* dimension in the usual left-to-right fashion and that they vote for the party closest to them. Moreover, preferences are *single-peaked* and *symmetric*, and voters can choose to abstain if they are too distant from a party. Under these conditions, and by allowing the distribution of voters to vary along the scale, Downs finds that Hotelling's conclusion that the parties in a two-party system inevitably converge is no longer necessarily true. If the distribution is approximately normal, parties will still move towards the median since they can attract more votes in the center than they would lose at the extremes due to abstention. If, on the other hand, the electorate is polarized, meaning that most voters occupy opposite sides of the distribution and very few can be found in the center, the two parties will tend to diverge toward the extremes and adopt very different ideologies. He concludes that the political systems' stability depends on the distribution of voters' preferences, which is a

variable in the long run.

> This reasoning implies that stable government in a two-party democracy requires a distribution of voters roughly approximating a normal curve. When such a distribution exists, the two parties come to resemble each other closely. Thus, when one replaces the other in office, no drastic policy changes occur, and most voters are located relatively close to the incumbent's position no matter which party is in power - (Downs, 1957).

The importance of the middle of the distribution, where the concept of *middle* in politics is captured by the *median* (Hinich and Munger, 1997), was also recognized at around the same time by Duncan Black, who derived the famous median voter theorem in *The Theory of Committees and Elections* (1958).

The spatial models of voting that have emerged in economics and political science are still based on the fundamental idea that voters are positioned along *ideological* dimensions. In such models, voters and parties are represented by points in an abstract low-dimensional space. Downs' *left-right* space is an example where there is one single ideological dimension. Each voter has a utility function over this space, which decreases with the distance between the position of the voter (the *ideal point*) and that of the party.[3] The crucial idea is the following. Citizens tend to have preferences on a wide range of issues (e.g., abortion, immigration, taxation), making the political space a complex, high-dimensional issue space where each issue has its own dimension. However, in practice "attitudes appear to be organized by positions along a small number of latent dimensions" (Lipovetsky, 2022). We commonly refer to these latent dimensions as *ideologies*. The classic *left-right* paradigm is an example.

The use of the terms *left* and *right* as a spatial metaphor in a political context dates back to just after the French Revolution of 1789. At first, these terms were used to describe the physical position of the groups that sat in the National Assembly. Over time, they became associated with the political preferences of the groups themselves. The ones on the left (Jacobins) were in favor of change, and those on the right (Girondins)

[3]There are many variants of the spatial theory of voting. In general, they can be differentiated between the classic Davis-Hinich-Ordeshook (Davis et al., 1970) variant, in which the latent dimensions are issues, and the neo-dowsonian (Enelow and Hinich, 1984) approach, where the latent dimensions are ideological. In this study, we only focus on the latter.

defended the status quo (Hinich and Munger, 1997). Nowadays, our interpretation of this dimension is not very far off. What makes this dimension *ideological* is that a voter's position on this scale is informative of her political preferences on various issues. In other words, it captures variation in the issue space.

The fundamental consequence of the existence of *ideologies* is that just a few underlying dimensions can explain the wide range of political preferences. This means it is possible to map individual positions in the complex issue space onto an underlying lower-dimensional *ideological* space, allowing us to model voting behavior with only a few latent dimensions. The answer to the question of *how many* and *which* ideological dimensions are needed to depict the issue space accurately depends on the political context. For example, in current American politics, a single *left-right* or *liberal-conservative* dimension may constrain political attitudes (Lipovetsky, 2022). In fact, "one of the underappreciated aspects of contemporary political polarization has been how a diverse set of policy conflicts — from abortion to gun control to immigration — have collapsed into the dominant economic liberal-conservative dimension of American politics" (Hare and Poole, 2013). Regarding the European political context, Bakker et al. (2012) shows with Confirmatory Factor Analysis (CFA) that three distinct dimensions are present on the *supply* side (i.e., in the competition among political parties) in most European countries. However, the results change if one looks at the *demand* side (i.e., the orientation of voters). Wheatley and Mendez (2021), for example, provides evidence that a three-dimensional model does not fit the data best and that "different bundles of issues group together and form dimensions in different ways in different countries". The results also depend on the statistical method employed to recover the dimensions underlying political preferences, the so-called *scaling procedure*. In general, successful scaling techniques need to be able to answer the following question: "How many latent dimensions of political difference do we need to describe and analyze the political problem at hand without destroying 'too much' information?" (Benoit and Laver, 2012).

This idea that preference variation in the issue space can be captured by a small number of latent dimensions or *ideologies* is consistent with Converse's (1964) belief system theory and his notion of *constraint*. This concept describes the tendency of individuals

to bundle many issue positions together as part of the same ideology. Therefore, one's ideology is informative of that person's stance on many issues at once. According to Downs (1957), the reason why voters bundle together policy preferences, or, in more general terms, the reason why ideologies exist, can be explained by the presence of imperfect information:

> In a complex society the cost in time alone of comparing all the ways in which the policies of competing parties differ is staggering. (...) if the voter discovers a correlation between each party's ideology and its policies, he can rationally vote by comparing ideologies rather than policies. (...) Thus, lack of information creates a demand for ideologies in the electorate.

Ideologies would not exist in a world where knowledge is perfect and information is costless. If citizens were aware of the exact parties' stances on all the issues they care about, they could choose which party to vote for by simply comparing them. Ideologies become helpful only when we assume that knowledge is imperfect and that information is costly. Under these conditions, ideologies can be used as proxies of the parties' differentiating stands, thus saving voters the cost of informing themselves on every single issue. The basic space theory thus posits that voters perceive parties as bundles of different issues represented by points in the abstract space formed by the relevant ideological dimensions. They then evaluate each platform by comparing it to their own *ideal point*, weighting each dimension based on its relative importance (or *salience*).

The foundation of the empirical application of the spatial theory of voting lies in random utility models (McFadden, 1974). It was Poole and Rosenthal who, for the first time, combined the spatial theory of voting and random utility models to study parliamentary roll call data. They developed NOMINATE (Poole and Rosenthal, 1985, 1991, 2000, 2011; Poole, 2005; McCarty et al., 2016), a successful multidimensional scaling method to measure the political ideology of the members of US Congress across time. They demonstrated that, despite its complexity, roll call voting can be modeled with just two dimensions: one is either the usual left-right dimension on economic issues or a liberal-conservative scale, and the other is related to salient social issues of the day.

With advances in computing power, more elaborate scaling techniques have been

developed to estimate an Euclidean map that places voters and parties in the same ideological space. Traditionally, these have been based on either roll call data or surveys (Jacoby and Armstrong II, 2014; Hare and Poole, 2018; Hare et al., 2018; Struthers et al., 2020), but scaling techniques that take advantage of digital trace data have also started to appear recently. Barberá (2015) was the first to show that it is possible to recover the liberal or conservative attitudes of millions of Twitter users in the US from their digital traces. Morales et al. (2022) then applied these network ideological scaling methods to the European context, showing that they work beyond one-dimensional opinion scales.

Many other studies have focused on estimating the features of voters' utility function (i.e., the *salience* of different dimensions) in the ideological space (Enelow and Hinich, 1985; Schofield et al., 1998; Dow, 1998; Quinn et al., 1999; Thurner and Eymann, 2000; McAllister et al., 2015; Magni-Berton and Panel, 2018; Stiers, 2022; Lucas et al., 2023). Recently, Danieli et al. (2022) has looked at the relevance of different factors - changes in party positions, voter attributes, and voter priorities (i.e., the salience of different dimensions) - in explaining the recent growth in electoral appeal of the populist radical right in Europe. Their findings indicate that voters attach increasing importance to the issues owned by these parties, which in turn explains their electoral success. Galasso et al. (2024) has also investigated the role of (dis)trust in political institutions on the support of populist parties, positing that "voters who no longer trust traditional parties either abstain or vote for populist parties committing to specific (economic or identity) policies, as long as the commitment credibility is strong enough". Voters' priorities have also changed in the US, with voters attaching increasing importance to cultural issues (Bonomi et al., 2021). Gennaioli and Tabellini (2023) investigate the role of *identity* in shaping voters' beliefs. Their findings indicate that shifting social identities are important drivers of changes in voter demands, explaining also why cultural groups have become more polarized on social policy and redistribution while opposite classes have become less polarized on the latter.

3 The spatial voting model formulated in the framework of random utility models

Starting with the standard spatial model, we assume that there are N voters in an economy and that all voters participate. Let R^G be the G-dimensional Euclidean space representing the ideological space. Points in the space represent voters and parties, and each voter is assumed to have a well-defined utility function over it. More precisely, each voter n is represented by a point $\boldsymbol{x_n} = [x_{n1}, \ldots, x_{nG}]' \in R^G$, where x_{ng} is the position of voter n on the ideological dimension g. This is the point of the maximum utility of the voter, and it is commonly referred to as her *ideal point* or *bliss point*. Each party j, instead, is represented by the point $\boldsymbol{a_j} = [a_{j1}, \ldots, a_{jG}]' \in R^G$, where a_{jg} is the ideological position of party j on dimension g. Voters share the same choice set $A = \{\boldsymbol{a_1}, \ldots, \boldsymbol{a_J}\}$, with $J \geq 2$, and they choose the alternative that provides the greatest utility. Therefore, voter n chooses party k if and only if $U_{nk} > U_{nj}\ \forall j \neq k$ (i.e., we assume *sincere voting*). Voters base their evaluation on each party's relative *weighted* distance from their ideal point, with the weights reflecting the *salience* of each dimension. Specifically, we assume

$$U_{nj} = \beta_j - \sum_{g=1}^{G} \beta_g \cdot d(x_{ng} - a_{jg}) + \epsilon_{nj}, \tag{1}$$

where $d(x_{ng} - a_{jg})$ is a measure of distance between the agent and the party on dimension g. β_g is a weighting constant that determines the salience of the g^{th} dimension. These weights represent how agents trade off closeness on one dimension against distance on another when evaluating different parties (Thurner, 2000). The weight's sign is positive when voters prefer parties close to them on a given dimension. The magnitude of the weights, in absolute values, indicates the relative importance of different dimensions for the voters. The larger the magnitude, the more important the dimension is. The standard assumption is that these weights are identical across all voters (i.e., the electorate is *homogeneous*). β_j is an alternative-specific constant; it captures the average effect of all factors not included in the model on utility from party j. We use it to capture the party's *valence*; i.e., factors such as the party leader's charisma or the party's reputation that can

affect voters' utility (see Danieli et al., 2022). Finally, ϵ_{nj} represents all the unobserved factors that can affect utility but are not included in the model. Let

$$V_{nj} = V(\boldsymbol{x_n}, \boldsymbol{a_j}) = \beta_j - \sum_{g=1}^{G} \beta_g \cdot d(x_{ng} - a_{jg}).$$

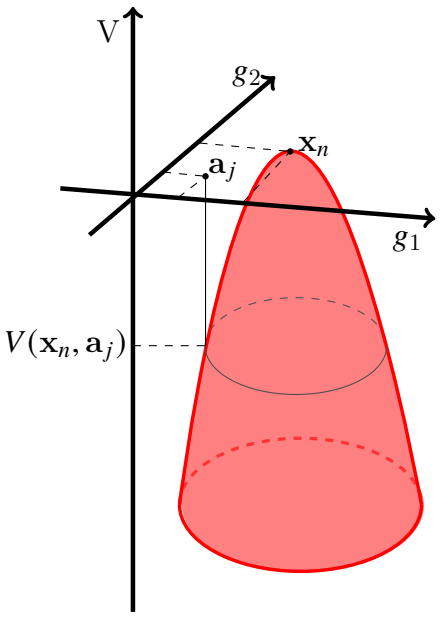
Defined in this way, the function V_{nj} is called the *representative utility* of the agent; it relates the observed attributes of the alternatives and the decision maker to the decision maker's utility. We can then rewrite equation (1) as

$$U_{nj} = V_{nj} + \epsilon_{nj}$$

It is important to note that the fact that $V_{nj} \neq U_{nj}$, meaning that utility is a random function, does not indicate a lack of information on the part of the decision maker. Instead, it suggests that there are aspects of utility that we, as researchers, do not observe (see Train, 2009).

Figure 1: Representative utility of an agent with position $\boldsymbol{x_n} = [6,\ 6]'$, assuming $\beta_1 = \beta_2 = 1$ and $\beta_j = 0\ \forall j$.

(a) $V(\boldsymbol{x_n}, \boldsymbol{a_j}) = -(x_{n1} - a_{j1})^2 - (x_{n2} - a_{j2})^2$



(b) Contour plot, view from top

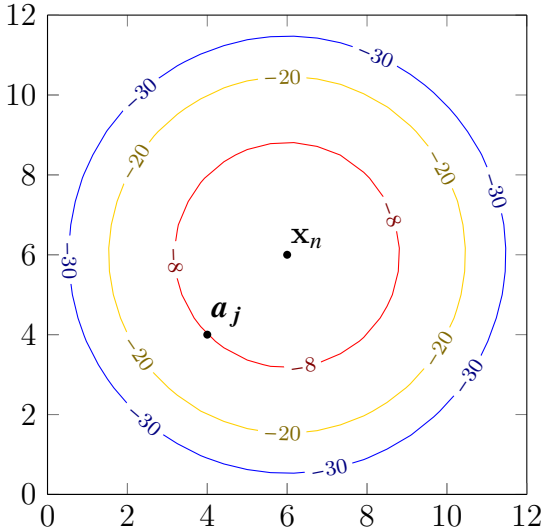


Figure 1 (a) provides a visual representation of the representative utility $V(\cdot)$ of an agent with position $\boldsymbol{x_n} = [6,\ 6]'$ assuming that: $G = 2$ (i.e. there are only two dimensions), $\beta_1 = \beta_2 = 1$ (i.e. both dimensions have the same salience), $\beta_j = 0\ \forall j$ (all parties have equal valence), and $d(x_{ng} - a_{jg}) = (x_{ng} - a_{jg})^2$ (i.e. standard Euclidean distance). Let g_1 and g_2 be the two dimensions of the ideological space. These define the g_1g_2-plane, the function $V(\cdot)$ then maps different combinations of g_1 and g_2 onto a third dimension (V-axis) which tells us the corresponding representative utility of the agent.

g_1 and g_2 could be any two dimensions. One can think, for example, of g_1 as an *economic issues* dimension and g_2 as a *social issues* one. The only assumptions that they need to satisfy are the following (see Hinich and Munger (1997)):

- *Ordering*: it must be possible to arrange parties and voters along each dimension, from *less* to *more*.
- *Continuity*: Between the positions of any two voters or parties lies another feasible position.

Given these assumptions, $V(\cdot)$ is shaped like a circular paraboloid pointing upwards below the g_1g_2-plane. This reflects *single-peaked* and *symmetric* preferences. Agents have one single point in the space that maximizes their utility, and as we move away from that point, their representative utility function slopes downward. The maximizer is the point $[6,\ 6]'$, which is exactly the agent's ideal point. For all other points $\boldsymbol{y}$ on the g_1g_2-plane $V(\boldsymbol{x_n}, \boldsymbol{y})$ is negative and decreasing with the distance between $\boldsymbol{x_n}$ and $\boldsymbol{y}$. Figure 1 (b) shows the *indifference curves*, these represent the sets of points in the two-dimensional g_1g_2-plane that give agent n the same level of (representative) utility. They capture the concept of indifference. When dimensions have equal salience (i.e., when $\beta_1 = \beta_2 = 1$), indifference curves correspond to all the points equidistant from $\boldsymbol{x_n}$; in other words, indifference curves are *circles*. For example, party j with position $\boldsymbol{a_j} = [4,\ 4]'$ would provide the agent with (representative) utility $V_{nj} = -(6-4)^2 - (6-4)^2 = -8$. Any party that lies on the same indifference curve as party $\boldsymbol{a_j}$ (the red circle in the Figure) would provide agent n with the same level of utility. Finally, the set of all alternatives inside the indifference curve on which $\boldsymbol{a_j}$ lies is called *preferred-to-set*; i.e., the set of all parties that are closer to agent n's ideal point and thus provide greater utility.

Figure 2: Indifference curves of an agent with position $\boldsymbol{x_n} = [6,\ 6]'$, assuming $\beta_1 = 2$ and $\beta_2 = 1$ and $\beta_j = 0\ \forall j$.

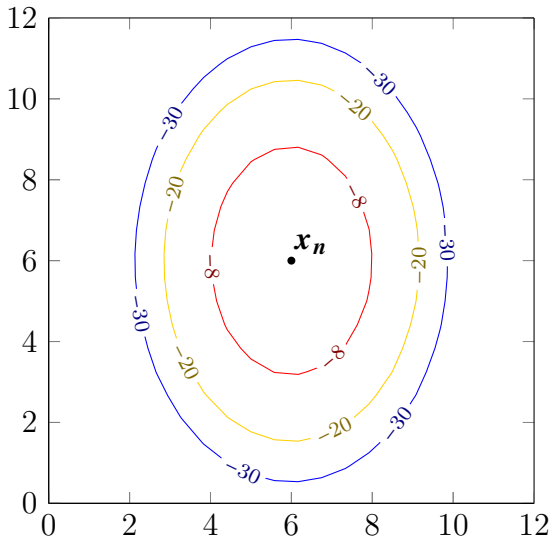


Figure 2 shows what happens to the indifference curves when the salience of the two ideological dimensions g_1 and g_2 is different (i.e., when $\beta_1 \neq \beta_2$). For example, if $\beta_1 = 2 > \beta_2 = 1$, agents will give twice more weight to dimension g_1 relative to dimension g_2. This means, for example, that if a party moves one step away from the agent's ideal point on dimension g_1, it would have to move two steps closer on dimension g_2 to remain on the same indifference curve. In general, indifference curves are "tall" when the horizontal dimension is more salient than the vertical one and "wide" in the opposite case (Hinich and Munger, 1997). Our goal in the first part of the study will be to empirically estimate the salience of each dimension based on the results of the 2022 national election in Italy.

3.1 Derivation of choice probabilities

Consider again equation (1) and let $z_{njg} = -d(x_{ng} - a_{jg})$. In other words, z_{njg} is the negative of the (observed) distance between agent n and party j on dimension g. This allows us to define the distance between the agent and the party on each dimension as a variable. In fact, "the structure of the multiattributive random utility model makes

it possible to treat policy-specific distances to each of the parties as attributes of these parties and to specify them as a generic variable" (Thurner and Eymann, 2000). Moreover, assume that $G = 2$ (i.e., the latent space is bi-dimensional) and that $d(x_{ng} - a_{jg}) = (x_{ng} - a_{jg})^2$ (i.e., standard Euclidean distance). Then,

$$U_{nj} = V_{nj} + \epsilon_{nj} = \beta_j + \beta_1 z_{nj1} + \beta_2 z_{nj2} + \epsilon_{nj} \tag{2}$$

Agent n will choose party k if the utility that she derives from this party is greater than the utility that she would derive from any other party (i.e., if $U_{nk} > U_{nj}\ \forall j \neq k$). It is easy to see then that the probability that agent n chooses party k is

$$\begin{aligned} P_{nk} &= Prob(U_{nk} > U_{nj};\ \forall j \neq k) \\ &= Prob(\beta_k + \beta_1 z_{nk1} + \beta_2 z_{nk2} + \epsilon_{nk} > \beta_j + \beta_1 z_{nj1} + \beta_2 z_{nj2} + \epsilon_{nj};\ \forall j \neq k) \\ &= Prob(\epsilon_{nj} - \epsilon_{nk} < (\beta_k - \beta_j) + \beta_1(z_{nk1} - z_{nj1}) + \beta_2(z_{nk2} - z_{nj2});\ \forall j \neq k) \\ &= Prob(\eta_{kj} < V_{nk} - V_{nj};\ \forall j \neq k) \end{aligned} \tag{3}$$

where $\eta_{kj} = \epsilon_{nj} - \epsilon_{nk}$. The functional form of P_{nk} depends on the assumption we make about the distribution of the error terms.

- ϵ_{nj} distributed as iid extreme value

Under this assumption, this becomes a standard discrete choice logit model. η_{kj} is distributed logistically (see Hartman, 1982) and the choice probabilities have the following closed-form expression (see Train, 2009):

$$P_{nk} = \frac{e^{V_{nk}}}{\sum_j e^{V_{nj}}} \quad \forall k \tag{4}$$

- $\boldsymbol{\epsilon_n}$ distributed as normal

Consider the vector composed of each ϵ_{nj}, labeled $\boldsymbol{\epsilon_n} = [\epsilon_{n1}, \ldots, \epsilon_{nJ}]'$. $\boldsymbol{\epsilon_n}$ is assumed to be jointly normal with a mean vector of zero and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1J} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{J1} & \sigma_{J2}^2 & \dots & \sigma_J^2 \end{bmatrix}.$$

Now let $\boldsymbol{\eta_k} = [\eta_{k1}, \dots, \eta_{kj}, \dots, \eta_{kJ}]' \ \forall j \neq k$, so that $\boldsymbol{\eta_k}$ has dimension $J-1$. Since the difference between two normals is normal, $\boldsymbol{\eta_k}$ is also jointly normally distributed with mean vector zero and covariance matrix

$$\Omega_k = \begin{bmatrix} \sigma_1^2 + \sigma_k^2 - 2\sigma_{1k} & \sigma_{12} - \sigma_{1k} - \sigma_{2k} + \sigma_k^2 & \dots & \sigma_{1J} - \sigma_{1k} - \sigma_{Jk} + \sigma_k^2 \\ \sigma_{21} - \sigma_{2k} - \sigma_{1k} + \sigma_k^2 & \sigma_2^2 + \sigma_k^2 - 2\sigma_{2k} & \dots & \sigma_{2J} - \sigma_{2k} - \sigma_{Jk} + \sigma_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{J1} - \sigma_{Jk} - \sigma_{1k} + \sigma_k^2 & \sigma_{J2} - \sigma_{Jk} - \sigma_{2k} + \sigma_k^2 & \dots & \sigma_J^2 + \sigma_k^2 - 2\sigma_{Jk} \end{bmatrix}$$

Therefore, the probability that voter n chooses party k becomes

$$P_{nk} = Prob(\eta_{kj} < V_{nk} - V_{nj}) = \int_{-\infty}^{V_{nk}-V_{n1}} \cdots \int_{-\infty}^{V_{nk}-V_{nJ}} \phi_k(\eta_{k1}, \dots, \eta_{kJ}) \, d\eta_{kJ} \dots d\eta_{k1} \ \ \forall j \neq k \tag{5}$$

where ϕ_k is a multivariate normal frequency with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Omega_k}$. Unfortunately, this integral has no closed form; it must be evaluated numerically through simulation.

3.2 Estimation

Let $\boldsymbol{\theta}$ be the vector of unknown parameters ($\beta_j, \beta_1, \beta_2$ and possibly the elements of Σ). With data on the individual choices of the agents in our sample, we can estimate $\boldsymbol{\theta}$ via MLE by maximizing the following likelihood function

$$L(\boldsymbol{\theta}) = \prod_{n=1}^{N} \prod_{k=1}^{J} P_{nk}^{y_{nk}} \tag{6}$$

where $y_{nk} = 1$ if agent n chooses party k and 0 otherwise.

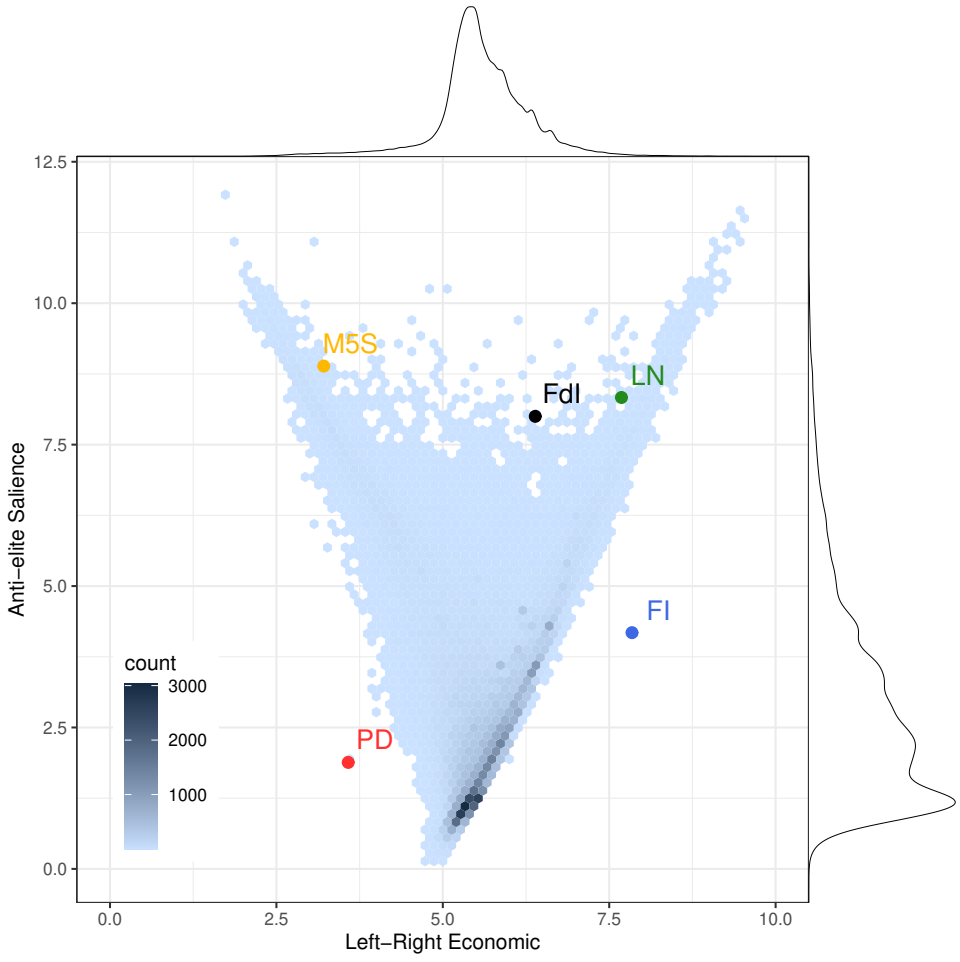
4 Data

The data we exploit in this study is obtained from Morales et al. (2022). To estimate the ideological positions of a sufficiently large Twitter population, Morales et al. (2022) utilizes a method known as ideological embedding. This methodology leverages online social network homophily (i.e., the assumption that people with similar attitudes follow each other online) to produce interpretable scales of positions for large numbers of users along dimensions of political issues and ideologies. This methodology was first developed by Barberá (2015), who managed to recover the *liberal-conservative* ideology of US Twitter users from their online network of followers and following. Morales et al. (2022) has successfully shown that this method works beyond one-dimensional opinion scales, making it more suitable for multidimensional European political settings. In practice, the method works as follows. First, the set of Italian Members of Parliament (MPs) present on Twitter and their followers is considered (filtering out inactive or bot accounts and considering only those that follow at least 3 MPs and have at least 25 followers)[4]. This resulted in 265,230 followers. Then, this set is represented as an adjacency matrix $A \in \{0,1\}^{|n.\ followers| \times |n.\ MPs|}$ (MPs are listed in columns and followers in rows, with values of 0 and 1 representing whether a user follows an MP), and Correspondence Analysis (CA) is applied to produce a lower-dimensional spatial representation. This operation generates a multidimensional latent position for each MP and the 265,230 followers in a low-dimensional space spanned by the PCs. These positions in space help explain how followers follow MPs based on proximity: the probability of one following the other is higher the closer they are. The results indicate that the first two PCs hold relatively more importance in explaining the topological network data. To understand to which political issues and ideologies these latent dimensions are linked, Chapel Hill Expert Survey (CHES) data (Jolly et al., 2022) is utilized. The CHES dataset includes party positions assessed by political science experts, who are asked to place European political parties on scales from 0 to 10 across 51 dimensions of political issues and ideologies. By comparing the positions of political parties according to the latent dimensions with their positions in the attitudinal CHES

[4]The data collection took place in the first half of 2020.

dimensions, Morales et al. (2022) found that the first two PCs are related to the *left-right economic* and *anti-elite salience* variables of the CHES, respectively. The first dimension is the classic left-right scale on economic issues. For example, an agent on the economic left may want the government to play a more active role in the economy than an agent on the right. On the other hand, the second dimension is more related to social issues, measuring the degree of anti-elitism. We can think of an agent that scores high on this dimension as having, for example, less trust in institutions relative to an agent with a lower score. The result is a dataset with the position of 265, 230 individuals in the ideological space spanned by these two dimensions. This is shown in Figure 3. Table 1 instead presents the descriptive statistics of the dataset.

Figure 3: Ideological space spanned by the *left-right economic* and *anti-elite salience* dimensions.



The Figure above shows the distribution of individuals and parties in the recovered

bi-dimensional ideological space. The color of each point indicates the number of Twitter users that can be found in that region of the space. The position of the parties instead is based on CHES. On the x-axis, we have the left-right economic dimension, while on the y-axis, we have the anti-elite salience one. The Figure shows that the distribution follows approximately a V-shape, with individuals more extreme on the left-right scale being relatively more anti-elite. Most individuals fall in the center-bottom of the distribution. This is consistent with what the marginal distributions show. While the left-right economic scale approximately follows a normal distribution, the other is strongly skewed to the right, meaning most individuals have a low score on the anti-elite salience dimension. Since CHES does not report minor parties' location on the two dimensions of interest for us, we will limit our analysis to the five major Italian parties: FdI, FI, LN, M5S, and PD.

Table 1: Descriptive statistics for *left-right economic* and *anti-elite salience* variables.

	left-right	*anti-elite*
count	265,230	265,230
mean	5.5881	2.9307
std	0.7247	1.8035
min	1.0395	0.0083
25%	5.2702	1.4721
50%	5.5329	2.4954
75%	5.9396	3.8580
max	12.1791	17.2923

Before we apply the spatial theory of voting to our dataset, it is essential to understand better the individuals included in it. Two primary challenges require attention when dealing with digital trace data. Firstly, our sample may be biased as Twitter users may not be representative of the entire population, especially in terms of age and gender. Secondly, the sample includes individuals who meet specific criteria (the ones we described previously) regardless of nationality, not just Italian Twitter users. Therefore, it is crucial to identify and isolate the subset of Italian users.

To address the first challenge, we aim to estimate the age and gender of individuals within our sample. This is essential to correct for potential sampling bias later on through post-stratification techniques. We use the *M3 system* developed by Wang et al. (2019b). This algorithm leverages data from the Twitter API, including users' profile pictures, bios, and screen names, to distinguish between individuals and organizations. The algorithm then assigns individuals to age categories ("<=18", "19-29", "30-39", and ">=40") and estimates their sex as *male* or *female*. Descriptive statistics for sex and age are presented in Table 2. Notably, the table reveals an over-representation of males (64.71%) compared to females.

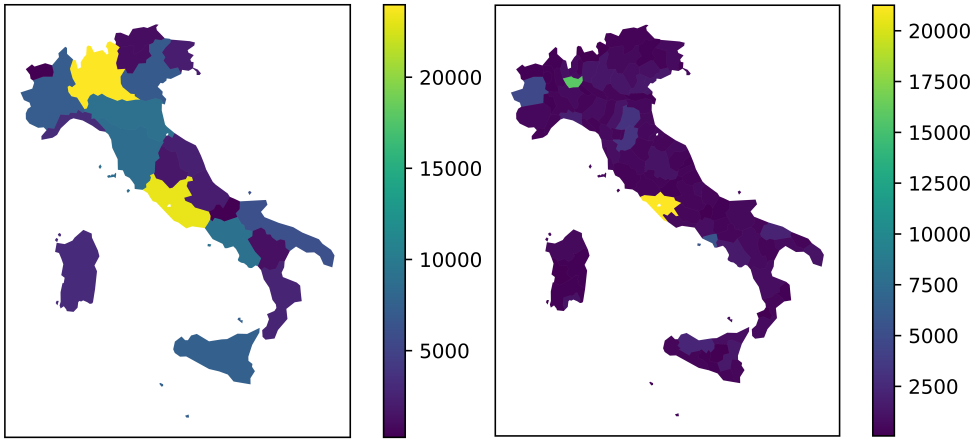
Table 2: Descriptive statistics for *sex*, *age* and *organization*.

	sex		*age*				*organization*	
	male	female	≤ 18	19-29	30-39	≥ 40	non-org	is-org
Freq.	113,437	61,876	6,010	9,062	25,359	51,608	183,637	27,891
Percent	64.71%	35.29%	6.53%	9.85%	27.55%	56.07%	86.81%	13.19%

To address the second issue, we exploit the location information provided by the users themselves. Twitter users can fill in their location details in the *location* field. We fetch this data from the Twitter API. Then, we match the user's self-reported location with the actual locations in Italy using a deterministic algorithm. This algorithm uses n-grams (i.e., groups of n adjacent words) present in the user's self-reported location to match them with real Italian locations, such as cities, provinces, regions, etc. These real locations are retrieved from the GeoNames database.[5] The distribution of Twitter users in our sample across regions and provinces, as well as the average score on each dimension at the province level, are shown in Figures 4 and 5.

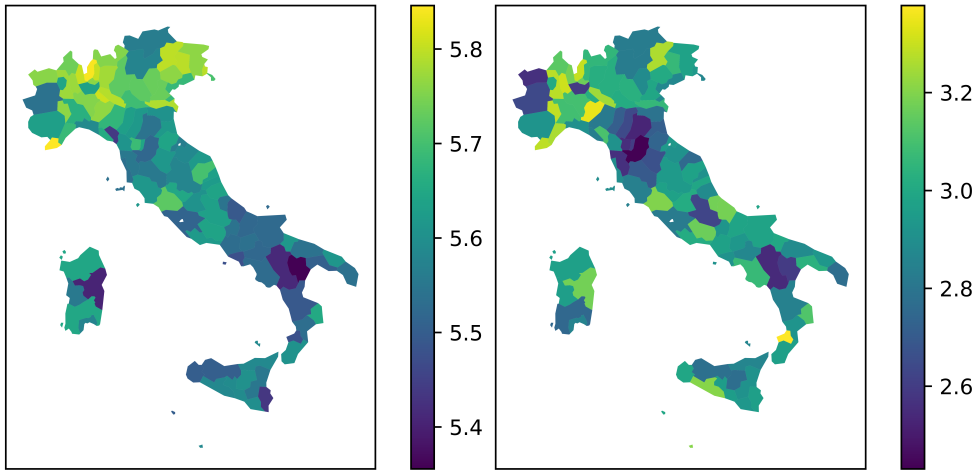
[5]The code, as well as a more detailed explanation of how it works, can be found at the following link: https://github.com/marvin-01/twitter_loctagger_it.

Figure 4: Number of users by region and province.



(a) By region (b) By province

Figure 5: Mean of *left-right* and *anti-elite* by province.



(a) *left-right* (b) *antielite*

As expected, most users are located in the two major Italian municipalities: Rome and Milan. Moreover, Northern provinces tend to have relatively higher scores on the left-right economic dimension than the South. On the other hand, the distinction is less clear regarding the anti-elite salience dimension. Finally, appendix A provides a validity test that shows that the resulting dataset (the ideological position of Twitter users from Morales et al. (2022) along with our demographic and location estimates) captures well the spatial distribution of the *left-right economic* and *anti-elite salience* ideological dimensions in Italy.

5 Mapping voters' preferences

We define the exact model specification as follows:

$$
\begin{aligned}
U_{nFdI} &= \beta_{FdI} + \beta_1 z_{nFdI1} + \beta_2 z_{nFdI2} + \epsilon_{nFdI} \\
U_{nFI} &= \beta_{FI} + \beta_1 z_{nFI1} + \beta_2 z_{nFI2} + \epsilon_{nFI} \\
U_{nLN} &= \beta_{LN} + \beta_1 z_{nLN1} + \beta_2 z_{nLN2} + \epsilon_{nLN} \\
U_{nM5S} &= \beta_{M5S} + \beta_1 z_{nM5S1} + \beta_2 z_{nM5S2} + \epsilon_{nM5S} \\
U_{nPD} &= \beta_{PD} + \beta_1 z_{nPD1} + \beta_2 z_{nPD2} + \epsilon_{nPD}
\end{aligned} \tag{7}
$$

This is the same as (1), where z_{nj1} and z_{nj2} are the distances (in negative terms) of agent n from party j on the *left-right economic* and *anti-elite salience* dimensions, respectively, $\forall j \in \{\text{FdI, FI, LN, M5S, PD}\}$. Before estimating the model, we must consider a crucial aspect of the behavioral decision-making process that affects the specification and estimation of any discrete choice model. This aspect is that "Only differences in utility matter" (Train, 2009).

Indeed, the absolute level of utility does not matter for the behavior of the decision maker: increasing U_{nj} by a constant k $\forall j$ would not change the choice of the decision-maker.[6] Therefore, only differences in the alternative-specific constants also matter. For instance, if the difference between β_{FdI} and β_{FI} is equal to d, increasing all alternative-specific constants by k would still generate the same difference of d in the two constants. This has repercussions also in terms of the estimation of the model.

Since there are infinite combinations of constants for which the differences are the same, it is impossible to estimate the constants themselves. Instead, the researcher needs to normalize the absolute level of the constants, which is commonly done by normalizing one of them to zero. In our case, we arbitrarily set the alternative-specific constant for the first choice, "Fratelli d'Italia", to zero (i.e., $\beta_{FdI} = 0$). Under this normalization, the constant for any other party β_j can be interpreted as the average effect of unincluded

[6] Note that the same is true from our perspective as researchers. This is clear by looking at equation (3); there we had that $P_{nk} = Prob(U_{nk} > U_{nj};\ \forall j) = Prob(U_{nk} - U_{nj} > 0;\ \forall j)$, which only depends on the difference in utility (see Train, 2009).

factors on the utility of j *relative* to FdI.

The major challenge in estimating this model with digital trace data is the lack of information on individual-level choices, which standard discrete choice modeling relies on for the estimation. Before showing how to estimate this model with aggregate data, we present a simple exercise that relies on our ability to retrieve users' political party preferences from the information they share online. This is intended to show how digital trace data alone can still provide valuable estimates of voters' preferences. However, since individual-level choices can be retrieved from online behavior only for a specific subset of users, the results of this estimation cannot be extended to the overall population.

5.1 Individual data

This approach is the simplest one, and it relies on the fact that individuals may reveal their political inclinations through their online activity. We are particularly interested in the information in a user's Twitter bio, as many people use this space to express their support for a political party. Figure 6 is an example of how users may easily reveal their political affiliation in their bio (in this case, the bio reads "I am a proud supporter of Lega and Italian"). Our approach involves using this information to determine a user's party preference and treating it as their choice in the context of a discrete choice model.

Figure 6: Example of a Twitter bio showing party support.

We exploit text analysis to identify which individuals in our sample express political preferences in their bio. We do so by matching specific keywords related to Italian parties (such as "PD" and "leghista"[7]) to the bios of all users. We identify a sub-sample of 1,686 individuals. We then go through the extracted sample and remove those that refer to a party negatively or neutrally and keep only the ones that show support (or affiliation) to

[7]The complete list of keywords can be found in Table 9 in appendix B.1

a party. We also remove those that show support for more than one party. This leaves us with a final sub-sample of $1,289$ individuals. Since these users will likely differ from the rest of the population, we will refer to them as the *politically active* ones. This subset will only include individuals who actively participate in politics or feel a strong enough connection to a party to communicate it to others actively. We can now estimate the model on this sub-sample using standard discrete choice modeling software.[8] The results are shown in Table 3.

Table 3: Logit model of spatial party choice of *politically active* voters (standard errors in parenthesis).

Variable	β	t-ratio	P-value
FI-constant	-0.341	-1.44	0.15
	(0.237)		
LN-constant	1.39^{***}	10.8	0.000
	(0.128)		
M5S-constant	1.67^{***}	4.55	0.000
	(0.367)		
PD-constant	-1.97^{***}	-5.71	0.000
	(0.345)		
Left-Right Economic	0.35^{***}	13	0.000
	(0.0269)		
Anti-elite Salience	0.244^{***}	10.6	0.000
	(0.0231)		
N. observations	1289		

The Table above shows that the coefficients on both the *left-right economic* and the *anti-elite salience* dimensions are positive and significant, which means that the utility that politically active agents receive from parties significantly decreases when the relative distance on each dimension increases. Furthermore, the magnitude of the weights is greater for the former dimension (0.35) than the latter one (0.244), meaning that politically active voters care relatively more about economic issues. Moreover, all the alternative specific constants are significant except for the one of Forza Italia. This means that no other factors (except the weighted distance on the two dimensions) affect voters' utility

[8] We used a software called Biogeme for Python; see Lancsar et al. (2017) for a review of standard statistical software packages that can be used to estimate DCMs.

from FI relative to FdI. In other words, politically active voters do not perceive significant differences across the two parties regarding *valence*. On the other hand, unobserved factors have a significantly positive average effect for LN and M5S and a significantly negative effect for PD.

The advantage of this approach is that it is straightforward to implement since the model can be estimated with standard libraries built for discrete choice modeling. However, it is essential to be cautious when interpreting these results. Since this approach relies on information individuals voluntarily express in their online profiles regarding their party preferences, extending these results to the broader population is impossible. People who express political preferences online will likely differ from everybody else, especially regarding party preferences. Still, we can gain valuable information regarding the preferences of those agents who are politically active.

To draw more general results, we need to find a way to take advantage of all the information contained in our sample. Since we do not have data on individual party preferences, we need to exploit estimation methods that rely on aggregate data. We propose an approach that relies on Maximum Likelihood Estimation (MLE). We estimate the model based on the results of the 2022 national election in Italy at the city level.

5.2 Aggregate data

5.2.1 Theoretical framework

Following Hartman (1982), we treat the use of aggregate data as a measurement error problem. Consider our model again, but now assume that we only observe average city estimates rather than individual values for each z_{njg}. In other words, for a voter n who lives in city c, what we observe is $\tilde{z}_{cjg} = z_{njg} - v_{cjg}$; i.e., the true distance of voter n from party j observed with an error. $\tilde{z}_{cjg}$ is the average distance (in negative terms) of all voters in city c from party j. Therefore, we have that

$$
\begin{aligned}
U_{nj} &= \beta_j + \beta_1 z_{nj1} + \beta_2 z_{nj2} + \epsilon_{nj} \\
&= \beta_j + \beta_1(\tilde{z}_{cj1} + v_{cj1}) + \beta_2(\tilde{z}_{cj2} + v_{cj2}) + \epsilon_{nj} \\
&= \beta_j + \beta_1 \tilde{z}_{cj1} + \beta_2 \tilde{z}_{cj2} + (\beta_1 v_{cj1} + \beta_2 v_{cj2} + \epsilon_{nj}) \\
&= \beta_j + \beta_1 \tilde{z}_{cj1} + \beta_2 \tilde{z}_{cj2} + \tilde{\epsilon}_{nj} \\
&= \tilde{V}_{nj} + \tilde{\epsilon}_{nj},
\end{aligned} \tag{8}
$$

where the new error term $\tilde{\epsilon}_{nj}$ derives from ϵ_{nj} as well as the measurement errors of z_{nj1} and z_{nj2}, generated by using aggregate data instead of individual-level data. The probability that voter n votes for party k then becomes

$$
\begin{aligned}
P_{nk} &= Prob(U_{nk} > U_{nj};\ \forall j \neq k) \\
&= Prob(\beta_k + \beta_1 \tilde{z}_{ck1} + \beta_2 \tilde{z}_{ck2} + \tilde{\epsilon}_{nk} > \beta_j + \beta_1 \tilde{z}_{cj1} + \beta_2 \tilde{z}_{cj2} + \tilde{\epsilon}_{nj};\ j \neq k) \\
&= Prob(\tilde{\epsilon}_{nj} - \tilde{\epsilon}_{nk} < \tilde{V}_{nk} - \tilde{V}_{nj};\ \forall j \neq k) \\
&= Prob(\tilde{\eta}_{kj} < \tilde{V}_{nk} - \tilde{V}_{nj};\ \forall j \neq k)
\end{aligned} \tag{9}
$$

where $\tilde{\eta}_{kj} = \tilde{\epsilon}_{nj} - \tilde{\epsilon}_{nk} = (\epsilon_{nj} - \epsilon_{nk}) + \beta_1(v_{cj1} - v_{ck1}) + \beta_2(v_{cj2} - v_{ck2})$.

Assuming that $\epsilon_{nj} \sim N(0, \sigma_j^2), v_{cj1} \sim N(0, \sigma_{v,j1}^2), v_{cj2} \sim N(0, \sigma_{v,j2}^2)$ and moreover assuming that all covariances among measurement errors (v_{cj1} and v_{cj2}) and between measurement errors and ϵ_{nj} are 0, we get that $\tilde{\eta}_{kj} \sim MN(0, \tilde{\Omega}_k)$. In particular (see appendix B.2),

- $Var(\tilde{\eta}_{kj}) = \sigma_j^2 + \sigma_k^2 - 2\sigma_{jk} + \sum_{g=1}^{2} \beta_g^2(\sigma_{v,jg}^2 + \sigma_{v,kg}^2) = \theta_{k,j}^2$
- $Cov(\tilde{\eta}_{kj}, \tilde{\eta}_{ki}) = \sigma_{ji} - \sigma_{jk} - \sigma_{ik} + \sigma_k^2 + \sum_{g=1}^{2} \beta_g^2 \sigma_{v,kg}^2 = \theta_{k,ji}$

and

$$
\tilde{\Omega}_k = \begin{bmatrix}
\theta_{k,1}^2 & \theta_{k,12} & \dots & \theta_{k,1J} \\
\theta_{k,21} & \theta_{k,2}^2 & \dots & \theta_{k,2J} \\
\vdots & \vdots & \ddots & \vdots \\
\theta_{k,J1} & \theta_{k,J2} & \dots & \theta_{k,J}^2
\end{bmatrix} \tag{10}
$$

Then, the probability that agent n votes for party j becomes

$$P_{nk} = Prob(\tilde{\eta}_{kj} < \tilde{V}_{nk} - \tilde{V}_{nj}) = \int_{-\infty}^{\tilde{V}_{nk}-\tilde{V}_{n1}} \cdots \int_{-\infty}^{\tilde{V}_{nk}-\tilde{V}_{nJ}} \tilde{\phi}_k(\tilde{\eta}_{k1}, \ldots, \tilde{\eta}_{kJ})\, d\tilde{\eta}_{kJ} \ldots d\tilde{\eta}_{k1}; \quad \forall j \neq k \tag{11}$$

Unfortunately, this integral has no closed form; we must evaluate it numerically through simulation. By letting once again $\boldsymbol{\theta}$ be the vector of unknown parameters of the model ($\beta_j, \beta_1, \beta_2$, and the elements of $\tilde{\Omega}_k$), we can estimate it via MLE by maximizing the likelihood function defined as follows:

$$L(\boldsymbol{\theta}) = \prod_{c \in C} \prod_{k=1}^{J} \prod_{n=1}^{M_{ck}} P_{nk}^{y_{nk}} = \prod_{c \in C} \prod_{k=1}^{J} P_{nk}^{M_{ck}} \tag{12}$$

where M_{ck} is the number of people in city c that voted for party k and C the set of cities. Note that P_{nk} will be equal for all voters in the same city c. This allows us to estimate the model's parameters without individual-level choice data. The log-likelihood is

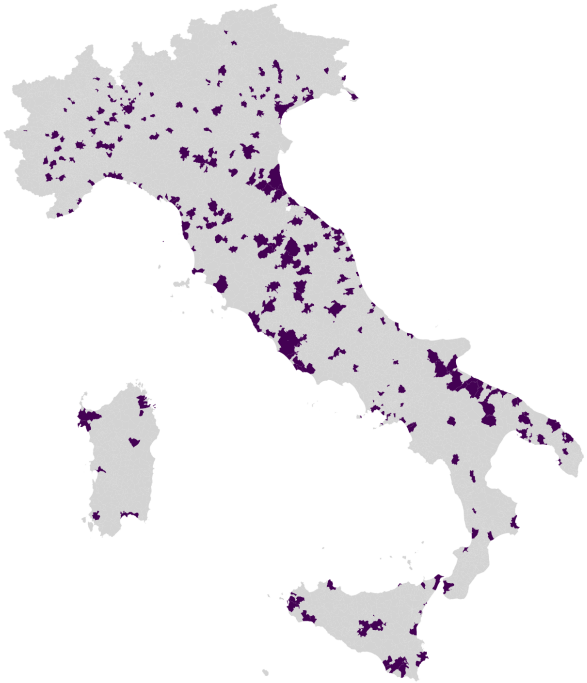
$$l(\boldsymbol{\theta}) = \sum_{c \in C} \sum_{k=1}^{J} M_{ck} ln(P_{nk}) \tag{13}$$

The intuition behind this approach is the following. Since we only have data on election results at the city level, we need to somehow transform the unit of analysis from the individual to the aggregate level. The idea is to construct a representative agent for each Italian municipality. The representative agent for city c is characterized by two attributes: $\tilde{z}_{cj1}$ and $\tilde{z}_{cj2}$. These are the estimates of the (perceived) distances of the representative agent of city c from party j on the two dimensions. They are constructed as the weighted average of the distances of all agents in city c from party j on each dimension (see appendix B.2 for a more detailed explanation). The uncertainty around these estimates ($\sigma_{v,nj1}$ and $\sigma_{v,nj2}$) depends on the variance of the distances of all the inhabitants of the city c from each party. Given our data, we can estimate the attributes of the representative agents of 266 Italian municipalities.[9] They range from relatively small ones, with around $5,034$ inhabitants in the voting age, to municipalities with more

[9]These municipalities are the ones for which we have enough observations from all the population strata to estimate the attributes of the representative agent accurately

than 2 million. The spatial distribution of these Italian municipalities is shown in Figure 7.

Figure 7: Spatial distribution of the cities included in the sample.



It is important to stress that, although we are aggregating the data, we are still taking into account the measurement error ($\sigma_{v,nj1}, \sigma_{v,nj2}\ \forall j$). This allows the resulting estimator to be consistent. If we used aggregate data ($\sigma_{v,nj1} \neq 0, \sigma_{v,nj2} \neq 0\ \forall j$) with standard probit techniques, we would ignore the presence of β_1^2, β_2^2 and the measurement error variances in the choice probabilities, and the resulting estimation of the model would be inconsistent.

5.2.2 Without abstention

To estimate the model, we start by making some simplifying assumptions. We assume that $\sigma_j^2 = \sigma^2$ and that $\sigma_{jk} = 0\ \forall j, k$. In other words, we assume that the unobserved factors have the same variance and are uncorrelated over alternatives. Hence, the error for one alternative provides no information about the error for another alternative. This assumption is appropriate if the utility is specified well enough that the remaining (unobserved) portion of utility is essentially "white noise" (Train, 2009). The full $\tilde{\Omega}_k$ (10) and P_{nk} (11) needed to be incorporated in the likelihood function (12) for $k \in \{$FdI, FI, LN,

M5S, PD} are formally developed in appendix B.2. The results are shown in Table 4.

Table 4: Probit model of spatial party choice based on the 2022 Italian nation election - without abstention (standard errors in parenthesis).

Variable	β
FI-constant	-0.752^{***}
	(0.004)
LN-constant	-0.403^{***}
	(0.002)
M5S-constant	0.518^{***}
	(0.003)
PD-constant	-0.069^{***}
	(0.004)
Left-Right Economic	0.099^{***}
	(0.000)
Anti-elite Salience	0.03^{***}
	(0.000)
N. observations	266

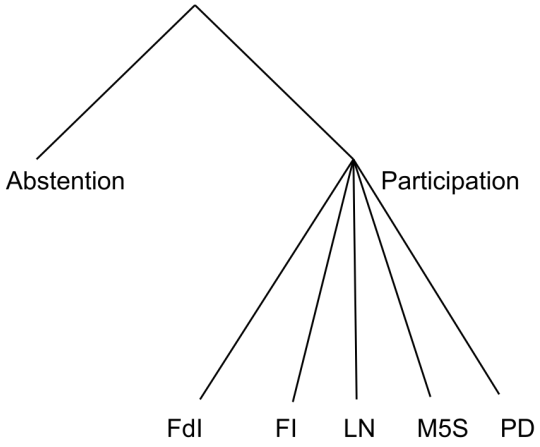
The Table above shows that the coefficients on both ideological dimensions are positive and significant. This means that Italian voters' utility from parties significantly decreases when the relative distance on each dimension increases. Moreover, the fact that the coefficient on the *left-right economic* dimension (0.09) is greater than that on the *antielite-salience* one (0.03) means that individuals care more about the former dimension than the latter. In terms of the alternative-specific constants, unlike for the politically active population, they are all significant. Furthermore, the signs of the constants indicate that unobserved factors have a positive effect only on the utility of M5S relative to FdI. Their impact is negative for all the other parties.

5.2.3 With abstention

So far, we have assumed that all the individuals in our sample vote; however, individuals may also decide to abstain. We model abstention under the assumption of expressive voting: each individual votes if and only if the maximum of the expected utilities from voting is greater than a fixed cost k of voting. This multi-stage decision problem can be

represented as the following decision tree.

Figure 8: Decision tree.



In the first step, the voter decides whether to abstain or participate. In the second, she evaluates the available parties and determines the best alternative. However, the two choices are not sequential; all the alternatives (five parties and the abstention option) are evaluated simultaneously. This is the case because the attributes of the lower branch alternatives, incorporated as expected maximum utility, influence the decision at the upper branch. Each voter evaluates the utility she would get by choosing any available party; if the utility she would get by choosing optimally is not greater than the cost of voting k, she will abstain. Under these conditions, the choice probabilities become

$$
\begin{aligned}
P_{An} &= Prob(max\{U_{nFdI}, U_{nFI}, U_{nLN}, U_{nM5S}, U_{nPD}\} \leq k) \\
&= Prob(U_{nFdI} \leq k \ \cap \ U_{nFI} \leq k \ \cap \ U_{nLN} \leq k \ \cap \ U_{nM5S} \leq k \ \cap \ U_{nPD} \leq k) \\
&= Prob(U_{nj} \leq k; \quad \forall j \in B) \\
&= Prob(\beta_j + \beta_1 \tilde{z}_{cj1} + \beta_2 \tilde{z}_{cj2} + \beta_1 v_{cj1} + \beta_2 v_{cj2} + \epsilon_{nj} \leq k; \quad \forall j \in B) \\
&= Prob(\tilde{\epsilon}_{nj} \leq k - \tilde{V}_{nj}; \quad \forall j \in B) \\
&= \int_{-\infty}^{k-\tilde{V}_{nFdI}} \cdots \int_{-\infty}^{k-\tilde{V}_{nPD}} \phi_{\tilde{\epsilon}_n}(\tilde{\epsilon}_{nFdI}, \cdots, \tilde{\epsilon}_{nPD}) \quad d\tilde{\epsilon}_{nPD}, \cdots, d\tilde{\epsilon}_{nFdI}
\end{aligned}
$$

where $\phi_{\tilde{\epsilon}_n}$ is distributed like a $N(\mathbf{0}, \begin{bmatrix} \theta_{FdI} & 0 & 0 & 0 & 0 \\ 0 & \theta_{FI} & 0 & 0 & 0 \\ 0 & 0 & \theta_{LN} & 0 & 0 \\ 0 & 0 & 0 & \theta_{M5S} & 0 \\ 0 & 0 & 0 & 0 & \theta_{PD} \end{bmatrix})$ and $\theta_j = \sigma^2 + \beta_1^2 \sigma_{v,nj1}^2 + \beta_2^2 \sigma_{v,nj2}^2$. The results are shown in Table 5.

Table 5: Probit model of spatial party choice based on the 2022 Italian nation election - with abstention (standard errors in parenthesis).

Variable	β
FI-constant	-0.511^{***}
	(0.004)
LN-constant	-0.479^{***}
	(0.002)
M5S-constant	0.289^{***}
	(0.003)
PD-constant	0.177^{***}
	(0.004)
Left-Right Economic	0.082^{***}
	(0.0004)
Anti-elite Salience	0.007^{***}
	(0.0002)
N. observations	266

Again, all coefficients are significant, and voters assign positive weights to the two ideological dimensions. However, adding the possibility for voters to abstain has two relevant effects. The first one is that the difference in magnitude between the salience of the two dimensions increases from 0.069 to 0.075, meaning that voters assign even more weight to the left-right economic dimension relative to the anti-elite salience one. The second effect is the coefficient for the constant for PD switches in sign (from -0.069 to 0.177). Finally, the cost of voting k is estimated to be 0.46 (significant at the 1% level).

Overall, our results indicate that the spatial theory of voting is appropriate to explain the results of the 2022 national election in Italy. Moreover, non-spatial, party-specific biases and the two ideological dimensions are significant explanatory variables. Regarding the relative importance of ideological dimensions for voters, we conclude that economic issues are still the most important.

6 Predicting individual voting behavior

This section investigates whether digital trace data can accurately predict individual voting behavior. We follow an approach similar to the one applied in supervised learning in the framework of spatial voting theory. *Supervised learning* is a classic data mining problem that aims to predict an output value associated with a specific input vector. The idea is to exploit a training set consisting of pairs of input vectors and output values to construct a predictor, which we can then use to predict output values for new input vectors. In our case, we want to construct a predictor that can predict the party choice of agents given their position in the ideological space recovered by Morales et al. (2022). More specifically, we define the input vector representing voter n as $\boldsymbol{x_n} = [z_{nFdI1}, z_{nFdI2}, \ldots, z_{nPD1}, z_{nPD2}]'$. In other words, each individual is represented by a vector of the relative distances between her and the parties on each dimension. The goal is to learn a predictor f that performs well on unseen test data drawn from the same source. In other words, "for a set of test input vectors $\{x_n\}$ with unknown output values $\{y_n\}$, we wish that $f(x_n) = y_n$ often" (Musicant et al., 2007b). Since the output data is nominal and there are more than two classes, this task is commonly referred to as *multi-class classification problem*. This particular problem at hand, however, has one relevant difference with respect to the classic one. Instead of having a training set with individual output values for each input vector, the output values are only available in aggregate across many input vectors. More precisely, instead of knowing the voting behavior of each individual in our sample, we only observe the output values (which, in our case, are the party shares) for multiple individuals at once (all those that live in the same city). This kind of problem is commonly referred to as an "aggregate output learning problem". Table 6 shows a sample dataset for this framework.

Table 6: Sample classification training and test sets. In the training set, classifications are known only in aggregate.

Left-Right Economic	Anti-elite Salience	City	Party Shares
5.5	2	Milan	FdI=33.7%, FI=9.5%, LN=12.9% M5S=12.6%, PD=31.1%
4	3	Milan	
5.2	3.5	Milan	
2.1	6	Milan	
1.8	1.2	Milan	
6.8	3.5	Rome	FdI=39.8%, FI=7.1%, LN=6.8%, M5S=19.1%, PD=27.3%
1.4	3.2	Rome	
3	2.7	Rome	
7	3.1	Rome	

Left-Right Economic	Anti-elite Salience	City	Party
5.1	2.3	Milan	?
7.2	3.4	Rome	?

More precisely, our training set consists of input vectors $\boldsymbol{x_{cn}}$, where c indicates to which city the input vector belongs and n is the specific input vector within that city. We assume that, for each city c, we observe the following set of aggregate output values: $s_c = \{s_{cFdI},\ s_{cFI},\ s_{cLN},\ s_{cM5S},\ s_{cPD}\}$, where s_{cj} is the share of votes received by party j in city c. Again, we consider the results of the 2022 national election in Italy. We wish to train a predictor f on this training data that can then operate on a single input vector to produce a single output value.

We specify the relationship between the input vector and the output value as the usual random utility model we have used so far, assuming all individuals participate. The utility that agent n derives from party j is defined as

$$\begin{aligned} U_{qnj} &= \beta_{qj} + \beta_{q1} z_{q,nj1} + \beta_{q2} z_{q,nj2} + \epsilon_{qnj} \\ &= V_{qnj} + \epsilon_{qnj} \end{aligned}$$

The only difference between this specification and the one used so far is that we allow for the possibility that the model parameters vary across different collections of

the input vectors. Each collection q represents a collection of cities. For example, we can group cities at the province or region level, allowing the model parameters to differ across different provinces or regions. If there is only one q (i.e., we do not define groups of cities), the specification falls back on (7). Moreover, instead of assuming that the error terms are normally distributed, we assume that ϵ_{qnj} is distributed i.i.d. extreme value. Empirically, the results would not change if we assumed that the error terms were independent and normally distributed with the same variance. However, this assumption speeds up the estimation process since the choice probabilities have a closed form, so there are no integrals to simulate. The probability that agent n in collection q votes for party k is

$$P_{qnk} = \frac{e^{V_{qnk}}}{\sum_j e^{V_{qnj}}};\ \forall k$$

6.1 Training algorithm

The algorithm we use to train the model is based on the Method of Simulated Moments (MSM). The idea is to minimize a certain distance between actual moments and simulated moments with respect to the vector of unknown parameters

$$\boldsymbol{\theta} = [\beta_{qFI}, \beta_{qLN}, \beta_{qM5S}, \beta_{qPD}, \beta_{q1}, \beta_{q2}]'\ \forall q$$

that generate the simulated data. We define the actual moments as the vote shares of each party in a city. More specifically, we define the MSM estimator as the solution to

$$\min_{\boldsymbol{\theta}}\ [\boldsymbol{s} - \hat{\boldsymbol{s}}(\boldsymbol{\theta})]' W [\boldsymbol{s} - \hat{\boldsymbol{s}}(\boldsymbol{\theta})] \tag{14}$$

where $\boldsymbol{s}$ is the vector of moments from the data and $\hat{\boldsymbol{s}}$ is the vector of simulated moments, which depends on $\boldsymbol{\theta}$. These contain the share of votes for each party in each city from actual and simulated data, respectively. Finally, W is the weighting matrix, where each weight is proportional to the size of the voter population of the corresponding city.

In general, a MSM algorithm works by iteratively (1) guessing the vector of param-

eters, (2) simulating the model based on the decision rules of the agents, (3) calculating moments from the simulated model (i.e., $\hat{\boldsymbol{s}}(\boldsymbol{\theta})$), and (4) comparing these to those from the data (i.e., $\boldsymbol{s}$). This procedure is repeated until the data and the model moments are as close as possible. This can be done, for example, by defining a grid of values for the vector of parameters based on the parameter space and iterating through all of them to find the minimizer (i.e., the *grid search method*). In practice, however, this method is impractical when the number of parameters rises beyond one or two. Instead, we rely on numerical methods and exploit a global optimization routine called Differential Evolution (DE) to find the values of the parameter vector that solve the minimization problem (14). This algorithm works by initializing a population of candidate solutions and then iteratively improving them by combining the current solutions; the best solution is chosen based on a condition. Conceptually, the process is the same as the grid search method, but it is not as time-consuming as the exhaustive search of the parameter space required by the latter. In practice, the simulation step (2) works as follows. After having set the vector of unknown parameters $\boldsymbol{\theta}$ to a specific value $\bar{\boldsymbol{\theta}}$, we simulate the choice that each individual would make given these parameters. In particular, for each agent n, we compute the probability $P_{qnj}(\bar{\boldsymbol{\theta}})$ that she votes for each party $j \in \{\text{FdI, FI, LN, M5S, PD}\}$. Assuming a representative sample, the percentage of people in each city c (which belongs to collection q) that vote for party j could be computed as

$$\hat{s}_{qcj}(\bar{\boldsymbol{\theta}}) = \frac{\sum_n P_{qnj}}{N_{qc}};$$

i.e., by taking the average of the individual probabilities of voting for party j across all agents that live in the city c. However, we need to account for sampling bias. Therefore, instead of computing the predicted share of votes for each party based on the equation above, we compute it as

$$\hat{s}_{qcj}(\bar{\boldsymbol{\theta}}) = \sum_{m \in M} \frac{N_{qcm}}{N_{qc}} \frac{\sum_n P_{qnj}}{N_{qcm}},$$

where M is the set of population categories[10] and N_{qcm} is the number of people in the city

[10] In our data, we have estimated the gender and age of each Twitter user. We have three categories for age ($19-29, 30-39, \geq 40$) and two categories for gender (male and female), therefore giving us six possible categories.

c (of collection q) that belong to the stratum m of the population. In other words, we first predict the percentage of votes party j would receive from agents in each category m of the population and then aggregate the results, giving a weight to each category that reflects their actual prevalence in the population.

6.2 Testing

We train the model based on the results of the 2022 national election in Italy. Since we do not know the voting choices of the individuals in our sample, to test the performance of this model in predicting individual-level party choices we rely on the sub-sample of the *politically active* individuals. We have already estimated the individual party preferences for these users in the previous section. The results are shown in Figure 9. We tested the model's performance for different sets of q, i.e., for different collections of cities. In particular, we tested the model's performance while allowing the model parameters to vary at the city or region level or not at all. In our case, the model had the best performance when the model parameters were not allowed to vary. Therefore, we only provide the results for this case.

Overall, the model has an accuracy (computed as the number of correct predictions over all the predictions) of 69.45%. The accuracy increases to 93.16% if we group all right-wing parties into the same category. This means that the model can classify correctly 93.16% of politically active users in either one of the following categories: M5S, PD, and Right-Wing coalition.

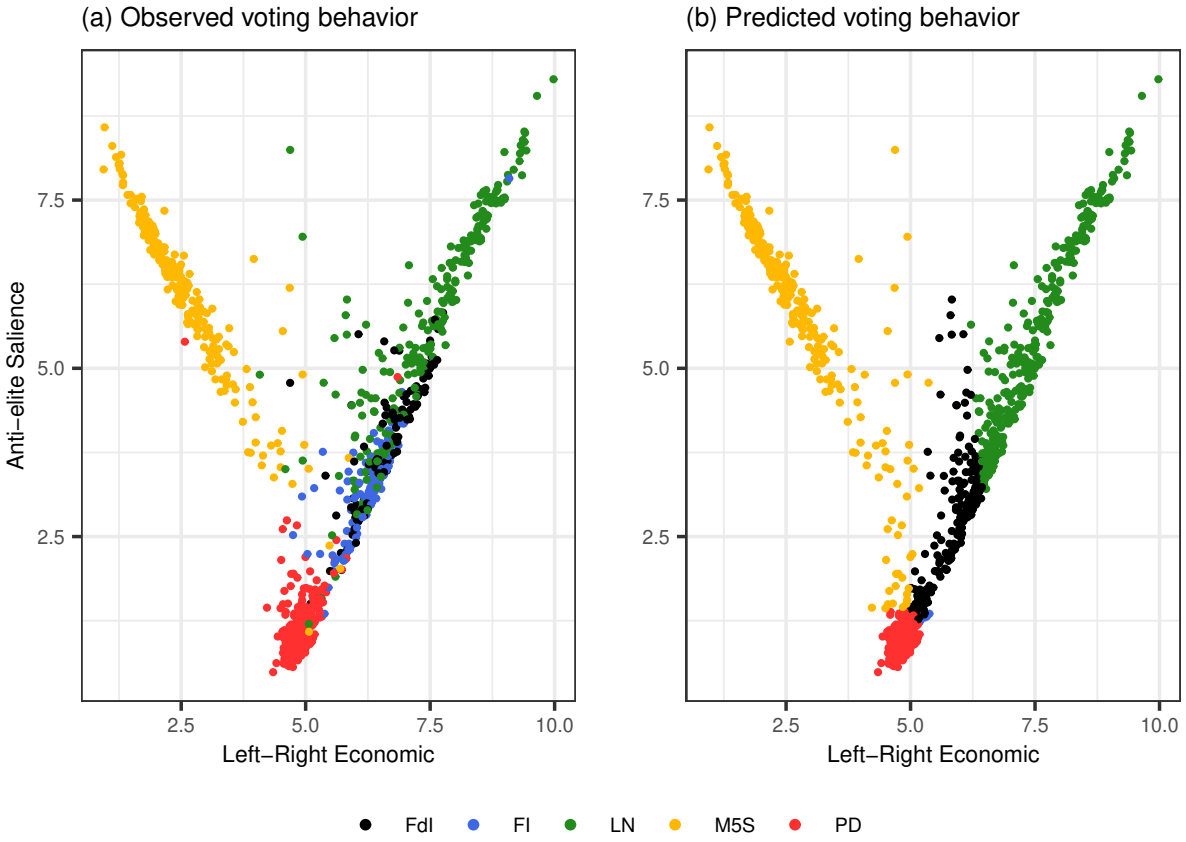
Table 7 instead reports the precision and recall by class. Precision for a given class in multi-class classification is computed as

$$Precision_j = \frac{TP_j}{TP_j + FP_j}$$

where TP_j is the number of true positives and FP_j the number of false positives of class j. Recall instead is computed as

$$Recall_j = \frac{TP_j}{TP_j + FN_j}$$

Figure 9: Observed and predicted choice of politically active voters.



where FN_j is the number of false negatives for class j. Precision refers to the fraction of instances where we correctly identified j out of all the cases where the algorithm declared j. On the other hand, recall refers to the fraction of instances where we correctly identified j out of all the cases where the true state of the world is j.

Table 7: Precision and recall by class.

Class	Precision	Recall
M5S	88.17%	98.3%
PD	99.19%	85.12%
RW coalition	91.52%	97.5%

7 Conclusion

This study aims to develop a framework to apply the spatial theory of voting to digital trace data. This is motivated by the limitations of traditional data collection methods in recovering citizens' attitudes and ideological positions, which are central to this theory. Recent advances in network ideological scaling methods have made it possible to recover individuals' positions on political issues and ideological scales from their digital traces. This data, however, comes with its challenges.

First, we have shown how to recover a sample of the Italian population starting from the data by Morales et al. (2022). The study was then divided into two parts. In the first part, we addressed the question: "With what probability will citizens with ideal points at x vote for one candidate, the other, or abstain?" (McKelvey, 1975). More specifically, we have demonstrated how to estimate a simultaneous model of voting and abstention with digital trace data, even without information on individual-level party choices. The empirical application was based on the results of the 2022 national election in Italy. Our results indicate that the spatial theory of voting is appropriate to explain the election results. Moreover, non-spatial, party-specific biases and the two ideological dimensions are important explanatory variables. Regarding the relative importance of the ideological dimensions for voters, economic issues are still the most important ones.

In the second part of the study, we focused on prediction. We have framed our problem as a multi-class classification problem with aggregate data and have shown how to predict the party choices of the electorate accurately. More precisely, we have trained a predictor on the data from Morales et al. (2022) and the results of the 2022 national election. Our results show that, given the ideological position of an individual, this predictor performs exceptionally well in predicting whether she will choose M5S, PD, or a party from the Right-Wing coalition (FdI, FI, or LN).

Appendices

A Validity test

To understand whether this dataset captures well the distribution of the *left-right* and *antielite* ideological dimensions in the Italian population, we can run a comparison with the results of the 2018 and 2022 elections. The idea is to see whether there is a correlation between the average value of the variables *left-right* and *antielite* in our dataset at the province level and the preferences expressed by voters at the ballot box in each province. We explain how this method works for the *left-right* dimension, but the same applies to the *antielite* one.

We can assign two scores for how left/right-wing each province is: one based on our Twitter users and the other based on the election results. If our dataset captures well the variation in this ideological dimension across provinces, we would expect these two measures to be correlated. Let's start by computing a measure that captures how left/right-wing a given province is based on the election results. We assign each party a score on the *left-right* dimension based on the corresponding variable in the Chapel Hill Expert Survey. Then, We assign a score to each province equal to the weighted average of the scores of the parties, where the weights are given by the percentage of votes obtained by each party in the given province. In other words, the score S on the ideological dimension $g \in \{\textit{left-right}, \textit{antielite}\}$ for province p is computed as follows:

$$S_p^g = \sum_j x_j^g \cdot \%votes_{jp}$$

where x_j^g is the score on the ideological dimension g for party j $\forall j = 1, \ldots, J$.

Then, we need a measure that captures how left/right-wing a given province is based on our Twitter users. For this, we could compute the average value of this variable by province. However, the result would not be representative of the actual population. Instead, to obtain a representative measure, we need to perform post-stratification.

Thus, for each province, we first compute the average value of the *left-right* dimension within each subgroup of the population[11]. Then, we aggregate these scores at the province level by weighting them by the percentage of people that belong to that subgroup in the given province. In other words, the score T on the ideological dimension $g \in \{left\text{-}right,\ antielite\}$ for province p is computed as follows:

$$T_p^g = \sum_{m \in M} \frac{N_{mp}}{N_p} \frac{\sum_n x_{np}^g}{N_{mp}}$$

where x_{np}^g is the score on the ideological dimension g for Twitter user n of province p, N_{mp} is the number of people that belong to subgroup m in province p, N_p is the total number of people in province p, and M the set of all population subgroups.

Table 8 reports the Pearson correlation coefficients of the correlation between the scores computed above at the province level for each ideological dimension. Namely, we are looking at the correlation between S^g and T^g, $\forall g \in \{left\text{-}right,\ antielite\}$.

Table 8: Pearson correlation coefficients.

	left-right	*antielite*
2018 Elections	0.7408	0.2821
2022 Elections	0.6981	0.3607

Based on these results, it seems that the indicator we constructed based on Twitter data is positively correlated with the indicator based on the election results on both dimensions. Moreover, the correlation is stronger for the $left-right$ dimension than the $antielite$ one.

[11] Since we have estimates for gender (male and female) and for three age groups above the voting age (19-29, 30-39, and $\geq$ 40) there are a total of six subgroups.

B Inference - full model statement and estimation

B.1 Individual data

Table 9: Keywords defining parties.

Party	Keywords
Fratelli d'Italia:	*Fratelli d'Italia*, *FdI*, *Meloni*
Forza Italia:	*Forza Italia*, *FI*, *Berlusconi*
Lega Nord:	*Lega Nord*, *LN*, *LSP*[1], *Salvini*
Movimento 5 Stelle:	*Movimento 5 Stelle*, *M5S*, *Conte*
Partito Democratico:	*Partito Democratico*, *PD*, *Letta*, *pdnetwork*, *dem*

[1] "Lega per Salvini Premier".

B.2 Aggregate data

For estimation we need P_{nj} and $\tilde{\Omega}_j$ for $j = 1, 2, 3, 4, 5$ in addition to the likelihood function (12). Recall that the measurement error variance $\sigma^2_{v,kg}$ $\forall g = 1, 2,$ $\forall k = 1, 2, 3, 4, 5$ is specific to each city. We avoid the subscript c to avoid making the notation too heavy.[12] However, this means that the covariance matrix $\tilde{\Omega}_j$ and in turn the functional form of P_{nj} will differ across cities. Given the assumption that $\sigma^2_j = \sigma^2$ and that $\sigma_{jk} = 0$ $\forall j, k$ we have that

$$
\begin{aligned}
Var(\tilde{\eta}_{kj}) &= Var(\tilde{\epsilon}_{nj} - \tilde{\epsilon}_{nk}) \\
&= Var(\epsilon_{nj} - \epsilon_{nk}) + \beta_1^2 Var(v_{cj1} - v_{ck1}) + \beta_2^2 Var(v_{cj2} - v_{ck2}) \\
&= 2\sigma^2 + \beta_1^2(\sigma^2_{v,j1} + \sigma^2_{v,k1}) + \beta_2^2(\sigma^2_{v,j2} + \sigma^2_{v,k2}) = \theta^2_{k,j}
\end{aligned}
$$

[12] We add the subscript back in the next section which regards the estimation.

$$
\begin{aligned}
Cov(\tilde{\eta}_{kj},\tilde{\eta}_{ki}) =& E(\tilde{\eta}_{kj}\cdot\tilde{\eta}_{ki}) - \underbrace{E(\tilde{\eta}_{kj})\cdot E(\tilde{\eta}_{ki})}_{0} \\
=& E\{[(\epsilon_{nj}-\epsilon_{nk})+\beta_1(v_{nj1}-v_{nk1})+\beta_2(v_{nj2}-v_{nk2})]\cdot \\
& [(\epsilon_{ni}-\epsilon_{nk})+\beta_1(v_{ni1}-v_{nk1})+\beta_2(v_{ni2}-v_{nk2})]\} \\
=& E\{[(\epsilon_{nj}-\epsilon_{nk})(\epsilon_{ni}-\epsilon_{nk})+\beta_1(\epsilon_{nj}-\epsilon_{nk})(v_{ni1}-v_{nk1})+\beta_2(\epsilon_{nj}-\epsilon_{nk})(v_{ni2}-v_{nk2})+ \\
& +\beta_1(v_{nj1}-v_{nk1})(\epsilon_{ni}-\epsilon_{nk})+\beta_1^2(v_{nj1}-v_{nk1})(v_{ni1}-v_{nk1})+\beta_1\beta_2(v_{nj1}-v_{nk1})(v_{ni2}-v_{nk2})+ \\
& +\beta_2(v_{nj2}-v_{nk2})(\epsilon_{ni}-\epsilon_{nk})+\beta_2\beta_2(v_{nj2}-v_{nk2})(v_{ni1}-v_{nk1})+\beta_2^2(v_{nj2}-v_{nk2})(v_{ni2}-v_{nk2})]\} \\
=& E[(\epsilon_{nj}-\epsilon_{nk})(\epsilon_{ni}-\epsilon_{nk})]+E[\beta_1^2(v_{nj1}-v_{nk1})(v_{ni1}-v_{nk1})]+E[\beta_2^2(v_{nj2}-v_{nk2})(v_{ni2}-v_{nk2})] \\
=& E[\epsilon_{nj}\epsilon_{ni}-\epsilon_{nj}\epsilon_{nk}-\epsilon_{nk}\epsilon_{ni}+\epsilon_{nk}^2]+\beta_1^2E[v_{nj1}v_{ni1}-v_{nj1}v_{nk1}-v_{nk1}v_{ni1}+v_{nk1}^2]+ \\
& +\beta_1^2E[v_{nj2}v_{ni2}-v_{nj2}v_{nk2}-v_{nk2}v_{ni2}+v_{nk2}^2] \\
=& VAR(\epsilon_{nk})+\beta_1^2VAR(v_{nk1})+\beta_2^2VAR(v_{nk2}) \\
=& \sigma^2+\beta_1^2\sigma_{v,k1}^2+\beta_2^2\sigma_{v,k2}^2
\end{aligned}
$$

$$
\tilde{\Omega}_1=\begin{bmatrix}\theta_{1,2}^2 & \theta_1 & \theta_1 & \theta_1\\ & \theta_{1,3}^2 & \theta_1 & \theta_1\\ & & \theta_{1,4}^2 & \theta_1\\ & & & \theta_{1,5}^2\end{bmatrix},\ \tilde{\Omega}_2=\begin{bmatrix}\theta_{2,1}^2 & \theta_2 & \theta_2 & \theta_2\\ & \theta_{2,3}^2 & \theta_2 & \theta_2\\ & & \theta_{2,4}^2 & \theta_2\\ & & & \theta_{2,5}^2\end{bmatrix},\ \tilde{\Omega}_3=\begin{bmatrix}\theta_{3,1}^2 & \theta_3 & \theta_3 & \theta_3\\ & \theta_{3,2}^2 & \theta_3 & \theta_3\\ & & \theta_{3,4}^2 & \theta_3\\ & & & \theta_{3,5}^2\end{bmatrix},
$$

$$
\tilde{\Omega}_4=\begin{bmatrix}\theta_{4,1}^2 & \theta_4 & \theta_4 & \theta_4\\ & \theta_{4,2}^2 & \theta_4 & \theta_4\\ & & \theta_{4,3}^2 & \theta_4\\ & & & \theta_{4,5}^2\end{bmatrix},\ \tilde{\Omega}_5=\begin{bmatrix}\theta_{5,1}^2 & \theta_5 & \theta_5 & \theta_5\\ & \theta_{5,2}^2 & \theta_5 & \theta_5\\ & & \theta_{5,3}^2 & \theta_5\\ & & & \theta_{5,4}^2\end{bmatrix}
$$

$$
\begin{aligned}
\tilde{\Omega}_1 &= \begin{bmatrix}2\sigma^2+\beta_1^2(\sigma_{v,21}^2+\sigma_{v,11}^2)+\beta_2^2(\sigma_{v,22}^2+\sigma_{v,12}^2) & \cdots & \sigma^2+\beta_1^2\sigma_{v,11}^2+\beta_2^2\sigma_{v,12}^2\\ \vdots & \ddots & \vdots\\ \sigma^2+\beta_1^2\sigma_{v,11}^2+\beta_2^2\sigma_{v,12}^2 & \cdots & 2\sigma^2+\beta_1^2(\sigma_{v,51}^2+\sigma_{v,11}^2)+\beta_2^2(\sigma_{v,52}^2+\sigma_{v,12}^2)\end{bmatrix} \\
&= \begin{bmatrix}\beta_1(\sigma_{v,21}^2+\sigma_{v,11}^2)+\beta_2^2(\sigma_{v,22}^2+\sigma_{v,12}^2) & \cdots & \beta_1\sigma_{v,11}^2+\beta_2^2\sigma_{v,12}^2\\ \vdots & \ddots & \vdots\\ \beta_1\sigma_{v,11}^2+\beta_2^2\sigma_{v,12}^2 & \cdots & \beta_1(\sigma_{v,51}^2+\sigma_{v,11}^2)+\beta_2^2(\sigma_{v,52}^2+\sigma_{v,12}^2)\end{bmatrix}+\sigma^2\begin{bmatrix}2 & \cdots & 1\\ \vdots & \ddots & \vdots\\ 1 & \cdots & 2\end{bmatrix}
\end{aligned}
$$

Then, for a given individual n with $k=1$ we have that

$$
P_{n1}=\int_{-\infty}^{\tilde{V}_{n1}-\tilde{V}_{n2}}\int_{-\infty}^{\tilde{V}_{n1}-\tilde{V}_{n3}}\int_{-\infty}^{\tilde{V}_{n1}-\tilde{V}_{n4}}\int_{-\infty}^{\tilde{V}_{n1}-\tilde{V}_{n5}}\tilde{\phi}_1(\tilde{\eta}_{12},\tilde{\eta}_{13},\tilde{\eta}_{14},\tilde{\eta}_{15})\ \ d\tilde{\eta}_{12}d\tilde{\eta}_{13}d\tilde{\eta}_{14}d\tilde{\eta}_{15} \quad (15)
$$

where $\tilde{\phi}_1$ is a multivariate normal with mean vector $\mathbf{0}$ and variance-covariance matrix $\tilde{\Omega}_1$.

Similar derivations hold for $P_{n2}, P_{n3}, P_{n4}, P_{n5}$.

Recall that $\tilde{V}_{nj} = \beta_1 \tilde{z}_{cj1} + \beta_2 \tilde{z}_{cj2}$ for individual n living in city c, where $\tilde{z}_{cjg}$ $\forall g$ is also specific to each city c and is computed as

$$\tilde{z}_{njg} = \sum_{m \in M} \frac{N_m}{N} \bar{z}_{mjg} \tag{16}$$

where $\bar{z}_{mjg} = \frac{\sum_n z_{njg}}{N_m}$ (i.e. the average distance on dimension g from party j), M is the set of population categories, N_m the number of people the belong to category m in city c and N the total number of people in city c. The parameters that need to be estimated are β_1 and β_2,; σ^2 is normalized to 1 and $\sigma^2_{v,jg}$ is replaced by its sample estimate $\hat{\sigma}^2_{v,jg}$ $\forall j = 1, \ldots, 5$ $\forall g$, which is computed as

$$\hat{\sigma}^2_{v,jg} = \sum_{m \in M} \frac{N_m}{N} \frac{\sum_n (z_{njg} - \bar{z}_{mjg})^2}{N_m - 1} \tag{17}$$

B.2.1 Model estimation

The log-likelihood function that we need to maximize is the following:

$$l(\boldsymbol{\theta}) = \sum_{c \in C} \sum_{k=1}^{5} M_{ck} ln(P_{ck}) = \sum_{k=1}^{5} \sum_{c \in C} M_{ck} ln(P_{ck}) \tag{18}$$

Start by defining the vectors $\boldsymbol{M} = \begin{pmatrix} M_{11} \\ \vdots \\ M_{c1} \\ \vdots \\ M_{15} \\ \vdots \\ M_{c5} \end{pmatrix}$, $\boldsymbol{P} = \begin{pmatrix} P_{11} \\ \vdots \\ P_{c1} \\ \vdots \\ P_{15} \\ \vdots \\ P_{c5} \end{pmatrix}$, where $M_{ck}$is the number of people in city c that voted for party k and P_{ck} is the probability that the representative agent from city c votes for party k. Therefore, equation (18) can be rewritten as

$$l(\boldsymbol{\theta}) = \boldsymbol{M'} \cdot (\boldsymbol{ln P_{ck}})_{\boldsymbol{ck}} \tag{19}$$

where

$$P_{ck} = \int_{-\infty}^{\tilde{V}_{ck}-\tilde{V}_{c1}} \cdots \int_{-\infty}^{\tilde{V}_{ck}-\tilde{V}_{c5}} \tilde{\phi}_{ck}(\tilde{\eta}_{ck1}, \cdots, \tilde{\eta}_{ck5})\, d\tilde{\eta}_{ck5} \cdots d\tilde{\eta}_{ck1} \tag{20}$$

and $\tilde{\phi}_{ck}$ is a multivariate normal with mean vector $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and variance-covariance matrix

$$\tilde{\Omega}_{ck} = \begin{bmatrix} \theta_{ck1}^2 & \theta_{ck} & \theta_{ck} & \theta_{ck} \\ & \ddots & \theta_{ck} & \theta_{ck} \\ & & \ddots & \theta_{ck} \\ & & & \theta_{ck5}^2 \end{bmatrix}.$$

Moreover,

- $\tilde{V}_{ck} = \beta_{0,k} + \beta_1 \tilde{z}_{ck1} + \beta_2 \tilde{z}_{ck2}$
- $\theta_{ck} = \sigma^2 + \beta_1 \sigma_{v,ck1}^2 + \beta_2 \sigma_{v,ck2}^2$
- $\theta_{ckj}^2 = 2\sigma^2 + \beta_1^2(\sigma_{v,ck1}^2 + \sigma_{v,cj1}^2) + \beta_2^2(\sigma_{v,ck2}^2 + \sigma_{v,cj2}^2)$

Now let

$$\tilde{\boldsymbol{V}} = \begin{pmatrix} \tilde{V}_{11} \\ \vdots \\ \tilde{V}_{c1} \\ \vdots \\ \tilde{V}_{15} \\ \vdots \\ \tilde{V}_{c5} \end{pmatrix} = \begin{pmatrix} \beta_{0,1} \\ \vdots \\ \beta_{0,1} \\ \vdots \\ \beta_{0,5} \\ \vdots \\ \beta_{0,5} \end{pmatrix} + \beta_1 \cdot \begin{pmatrix} \tilde{z}_{111} \\ \vdots \\ \tilde{z_{c}11} \\ \vdots \\ \tilde{z}_{151} \\ \vdots \\ \tilde{z}_{c51} \end{pmatrix} + \beta_2 \cdot \begin{pmatrix} \tilde{z}_{112} \\ \vdots \\ \tilde{z_{c}12} \\ \vdots \\ \tilde{z}_{152} \\ \vdots \\ \tilde{z}_{c52} \end{pmatrix} \tag{21}$$

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{11} \\ \vdots \\ \theta_{c1} \\ \vdots \\ \theta_{15} \\ \vdots \\ \theta_{c5} \end{pmatrix} = \sigma^2 + \beta_1 \cdot \begin{pmatrix} \sigma_{v,111}^2 \\ \vdots \\ \sigma_{v,c11}^2 \\ \vdots \\ \sigma_{v,151}^2 \\ \vdots \\ \sigma_{v,c51}^2 \end{pmatrix} + \beta_2 \cdot \begin{pmatrix} \sigma_{v,112}^2 \\ \vdots \\ \sigma_{v,c12}^2 \\ \vdots \\ \sigma_{v,152}^2 \\ \vdots \\ \sigma_{v,c52}^2 \end{pmatrix} \tag{22}$$

$$
\Theta = \begin{pmatrix}
\theta^2_{112} & \theta^2_{113} & \theta^2_{114} & \theta^2_{115} \\
\vdots & \vdots & \vdots & \vdots \\
\theta^2_{c12} & \theta^2_{c13} & \theta^2_{c14} & \theta^2_{c15} \\
\vdots & \vdots & \vdots & \vdots \\
\theta^2_{151} & \theta^2_{152} & \theta^2_{153} & \theta^2_{154} \\
\vdots & \vdots & \vdots & \vdots \\
\theta^2_{c51} & \theta^2_{c52} & \theta^2_{c53} & \theta^2_{c54}
\end{pmatrix}
$$

$$
\begin{aligned}
=& 2\sigma^2 + \beta_1^2 \cdot \left[\begin{pmatrix}
\sigma^2_{111} & \sigma^2_{111} & \sigma^2_{111} & \sigma^2_{111} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c11} & \sigma^2_{c11} & \sigma^2_{c11} & \sigma^2_{c11} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{151} & \sigma^2_{151} & \sigma^2_{151} & \sigma^2_{151} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c51} & \sigma^2_{c51} & \sigma^2_{c51} & \sigma^2_{c51}
\end{pmatrix} + \begin{pmatrix}
\sigma^2_{121} & \sigma^2_{131} & \sigma^2_{141} & \sigma^2_{151} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c21} & \sigma^2_{c31} & \sigma^2_{c41} & \sigma^2_{c51} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{111} & \sigma^2_{121} & \sigma^2_{131} & \sigma^2_{141} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c11} & \sigma^2_{c21} & \sigma^2_{c31} & \sigma^2_{c41}
\end{pmatrix} \right] \\
&+ \beta_2^2 \cdot \left[\begin{pmatrix}
\sigma^2_{112} & \sigma^2_{112} & \sigma^2_{112} & \sigma^2_{112} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c12} & \sigma^2_{c12} & \sigma^2_{c12} & \sigma^2_{c12} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{152} & \sigma^2_{152} & \sigma^2_{152} & \sigma^2_{152} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c52} & \sigma^2_{c52} & \sigma^2_{c52} & \sigma^2_{c52}
\end{pmatrix} + \begin{pmatrix}
\sigma^2_{122} & \sigma^2_{132} & \sigma^2_{142} & \sigma^2_{152} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c22} & \sigma^2_{c32} & \sigma^2_{c42} & \sigma^2_{c52} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{112} & \sigma^2_{122} & \sigma^2_{132} & \sigma^2_{142} \\
\vdots & \vdots & \vdots & \vdots \\
\sigma^2_{c12} & \sigma^2_{c22} & \sigma^2_{c32} & \sigma^2_{c42}
\end{pmatrix} \right]
\end{aligned} \tag{23}
$$

$$
\begin{aligned}
A &= \begin{pmatrix}
\tilde{V}_{11}-\tilde{V}_{12} & \tilde{V}_{11}-\tilde{V}_{13} & \tilde{V}_{11}-\tilde{V}_{14} & \tilde{V}_{11}-\tilde{V}_{15} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{V}_{c1}-\tilde{V}_{c2} & \tilde{V}_{c1}-\tilde{V}_{c3} & \tilde{V}_{c1}-\tilde{V}_{c4} & \tilde{V}_{c1}-\tilde{V}_{c5} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{V}_{15}-\tilde{V}_{11} & \tilde{V}_{15}-\tilde{V}_{12} & \tilde{V}_{15}-\tilde{V}_{13} & \tilde{V}_{15}-\tilde{V}_{14} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{V}_{c5}-\tilde{V}_{c1} & \tilde{V}_{c5}-\tilde{V}_{c2} & \tilde{V}_{c5}-\tilde{V}_{c3} & \tilde{V}_{c5}-\tilde{V}_{c4}
\end{pmatrix} \\
&= \begin{pmatrix}
\tilde{V}_{11} & \tilde{V}_{11} & \tilde{V}_{11} & \tilde{V}_{11} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{V}_{c1} & \tilde{V}_{c1} & \tilde{V}_{c1} & \tilde{V}_{c1} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{V}_{15} & \tilde{V}_{15} & \tilde{V}_{15} & \tilde{V}_{15} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{V}_{c5} & \tilde{V}_{c5} & \tilde{V}_{c5} & \tilde{V}_{c5}
\end{pmatrix} - \begin{pmatrix}
\beta_{0,2} & \beta_{0,3} & \beta_{0,4} & \beta_{0,5} \\
\vdots & \vdots & \vdots & \vdots \\
\beta_{0,2} & \beta_{0,3} & \beta_{0,4} & \beta_{0,5} \\
\vdots & \vdots & \vdots & \vdots \\
\beta_{0,1} & \beta_{0,2} & \beta_{0,3} & \beta_{0,4} \\
\vdots & \vdots & \vdots & \vdots \\
\beta_{0,1} & \beta_{0,2} & \beta_{0,3} & \beta_{0,4}
\end{pmatrix} - \\
&\quad -\beta_1 \cdot \begin{pmatrix}
\tilde{z}_{121} & \tilde{z}_{131} & \tilde{z}_{141} & \tilde{z}_{151} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{z}_{c21} & \tilde{z}_{c31} & \tilde{z}_{c41} & \tilde{z}_{c151} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{z}_{111} & \tilde{z}_{121} & \tilde{z}_{131} & \tilde{z}_{141} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{z}_{c11} & \tilde{z}_{c21} & \tilde{z}_{c31} & \tilde{z}_{c41}
\end{pmatrix} - \beta_2 \cdot \begin{pmatrix}
\tilde{z}_{122} & \tilde{z}_{132} & \tilde{z}_{142} & \tilde{z}_{152} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{z}_{c22} & \tilde{z}_{c32} & \tilde{z}_{c42} & \tilde{z}_{c152} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{z}_{112} & \tilde{z}_{122} & \tilde{z}_{132} & \tilde{z}_{142} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{z}_{c12} & \tilde{z}_{c22} & \tilde{z}_{c32} & \tilde{z}_{c42}
\end{pmatrix}
\end{aligned} \tag{24}
$$

Finally, we can compute $P_{ck}\ \forall c, k$ as

$$
P_{ck} = \tilde{\Phi}_{ck}(A_{c1}, \cdots, A_{c4}) \tag{25}
$$

where $\tilde{\Phi}_{ck}$is the Cumulative Distribution Function of a Multivariate Normal with mean vector $\mathbf{0}$ and variance-covariance matrix $\tilde{\Omega}_{ck}$.

C Prediction - full model statement and estimation

Let s_{cj} be the share of votes that party j received in the city c. The corresponding share of votes predicted by the model is

$$\hat{s}_{cj}(\bar{\boldsymbol{\theta}}) = \sum_{m \in M} \frac{N_{cm}}{N_c} \frac{\sum_n P_{nj}}{N_{cm}},$$

where N_c is the number of people in city c. Let $\boldsymbol{s}$ be the vector of data moment and $\hat{\boldsymbol{s}}$ the vector of simulated moments. We can define $\boldsymbol{s} = \begin{pmatrix} s_{11} \\ \vdots \\ s_{c1} \\ \vdots \\ s_{15} \\ \vdots \\ s_{c5} \end{pmatrix}$ and $\hat{\boldsymbol{s}} = \begin{pmatrix} \hat{s}_{11} \\ \vdots \\ \hat{s}_{c1} \\ \vdots \\ \hat{s}_{15} \\ \vdots \\ \hat{s}_{c5} \end{pmatrix}$. Moreover, let the weighting matrix $W = \begin{bmatrix} w_{11} & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & w_{c1} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & w_{15} & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & w_{c5} \end{bmatrix}$ where all the off-main diagonal elements are 0 and $w_{cj} = \frac{Nc}{N} \cdot \frac{1}{5} \ \ \forall c, j$, with N being the size of the Italian voting population.

References

R. Bakker, S. Jolly, and J. Polk. Complexity in the european party space: Exploring dimensionality with experts. *European Union Politics*, 13(2):219–245, 2012.

P. Barberá. Birds of the same feather tweet together. bayesian ideal point estimation using twitter data. 2015.

K. Benoit and M. Laver. The dimensionality of political space: Epistemological and methodological considerations. *European Union Politics*, 13(2):194–218, 2012.

D. Black et al. The theory of committees and elections. 1958.

G. Bonomi, N. Gennaioli, and G. Tabellini. Identity, Beliefs, and Political Conflict*. *The Quarterly Journal of Economics*, 136(4):2371–2411, 09 2021. ISSN 0033-5533. doi: 10.1093/qje/qjab034. URL https://doi.org/10.1093/qje/qjab034.

P. E. Converse. The nature of belief systems in mass publics (1964). *Critical review*, 18 (1-3):1–74, 2006.

O. Danieli, N. Gidron, S. Kikuchi, and R. Levy. Decomposing the rise of the populist radical right. 2022. ISSN 1556-5068. doi: 10.2139/ssrn.4255937. URL https://www.ssrn.com/abstract=4255937.

O. A. Davis, M. J. Hinich, and P. C. Ordeshook. An expository development of a mathematical model of the electoral process. *American political science review*, 64(2):426–448, 1970.

J. K. Dow. A spatial analysis of the 1989 chilean presidential election. 17 (1):61–76, 1998. ISSN 0261-3794. doi: 10.1016/S0261-3794(97)00050-4. URL https://www.sciencedirect.com/science/article/pii/S0261379497000504.

A. Downs. An economic theory of political action in a democracy. *Journal of political economy*, 65(2):135–150, 1957.

J. M. Enelow and M. J. Hinich. *The spatial theory of voting: An introduction*. CUP Archive, 1984.

J. M. Enelow and M. J. Hinich. Estimating the parameters of a spatial model of elections: An empirical test based on the 1980 national election study. *Political Methodology*, 11 (3/4):249–268, 1985. ISSN 01622021. URL http://www.jstor.org/stable/41289343.

V. Galasso, M. Morelli, T. Nannicini, and P. Stanig. The populist dynamic: Experimental evidence on the effects of countering populism. 2024. ISSN 1556-5068. doi: 10.2139/ssrn.4727221. URL https://www.ssrn.com/abstract=4727221.

N. Gennaioli and G. Tabellini. Identity politics, 2023. URL https://papers.ssrn.com/abstract=4410239.

C. Hare and K. T. Poole. The polarization of contemporary american politics, 2013. URL https://papers.ssrn.com/abstract=3363227.

C. Hare and K. T. Poole. Psychometric methods in political science. In *The Wiley Handbook of Psychometric Testing*, pages 901–931. John Wiley & Sons, Ltd, 2018. ISBN 978-1-118-48977-2. doi: 10.1002/9781118489772.ch28. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118489772.ch28. Section: 28 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118489772.ch28.

C. Hare, T.-P. Liu, and R. N. Lupton. What ordered optimal classification reveals about ideological structure, cleavages, and polarization in the american mass public. 176(1):57–78, 2018. ISSN 1573-7101. doi: 10.1007/s11127-018-0540-6. URL https://doi.org/10.1007/s11127-018-0540-6.

R. S. Hartman. A note on the use of aggregate data in individual choice models: Discrete consumer choice among alternative fuels for residential appliances. *Journal of Econometrics*, 18(3):313–335, 1982.

M. J. Hinich and M. C. Munger. *Analytical politics*. Cambridge University Press, 1997.

H. Hotelling. Stability in competition. *Economic Journal*, 39(153):41–57, 1929.

W. G. Jacoby and D. A. Armstrong II. Bootstrap confidence regions for multidimensional scaling solutions. 58(1):264–278, 2014. ISSN 1540-5907. doi: 10.1111/ajps.12056. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12056. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12056.

S. Jolly, R. Bakker, L. Hooghe, G. Marks, J. Polk, J. Rovny, M. Steenbergen, and M. A. Vachudova. Chapel hill expert survey trend file, 1999–2019. *Electoral studies*, 75: 102420, 2022.

E. Lancsar, D. G. Fiebig, and A. R. Hole. Discrete choice experiments: a guide to model specification, estimation and software. *Pharmacoeconomics*, 35:697–716, 2017.

S. Lipovetsky. Analyzing spatial models of choice and judgment, second edition: by david a. armstrong II, ryan bakker, royce carroll, christopher hare, keith t. poole, and howard rosenthal. CRC press. taylor & francis group, chapman and hall, boca raton, FL, 2021, ISBN 978-1-138-71533-2, 320 pp., $63.96 (hbk). 64(1):139–143, 2022. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.2021.2020519. URL https://www.tandfonline.com/doi/full/10.1080/00401706.2021.2020519.

J. Lucas, R. M. McGregor, and A. Bridgman. Spatial voting in non-partisan cities: A case study. 82:102599, 2023. ISSN 0261-3794. doi: 10.1016/j.electstud.2023.102599. URL https://www.sciencedirect.com/science/article/pii/S0261379423000215.

R. Magni-Berton and S. Panel. Manifestos and public opinion: testing the relevance of spatial models to explain salience choices. 16(5):783–804, 2018. ISSN 1740-388X. doi: 10.1057/s41295-017-0101-2. URL https://doi.org/10.1057/s41295-017-0101-2.

I. McAllister, J. Sheppard, and C. Bean. Valence and spatial explanations for voting in the 2013 australian election. *Australian Journal of Political Science*, 50(2):330–346, 2015.

N. McCarty, K. T. Poole, and H. Rosenthal. *Polarized America: The dance of ideology and unequal riches*. mit Press, 2016.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Fontiers in Econometrics*, pages 105–142. Academic press, New York, 1974.

R. D. McKelvey. Policy related voting and electoral equilibrium. *Econometrica: Journal of the Econometric Society*, pages 815–843, 1975.

P. R. Morales, J.-P. Cointet, and G. M. n. Zolotoochin. Unfolding the dimensionality structure of social networks in ideological embeddings. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '21, page 333–338, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391283. doi: 10.1145/3487351.3489441. URL https://doi.org/10.1145/3487351.3489441.

D. R. Musicant, J. M. Christensen, and J. F. Olson. Supervised learning by training on aggregate outputs. In *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, ICDM '07, page 252–261, USA, 2007a. IEEE Computer Society. ISBN 0769530184. doi: 10.1109/ICDM.2007.50. URL https://doi.org/10.1109/ICDM.2007.50.

D. R. Musicant, J. M. Christensen, and J. F. Olson. Supervised learning by training on aggregate outputs. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 252–261. IEEE, 2007b.

K. T. Poole. *Spatial models of parliamentary voting*. Cambridge University Press, 2005.

K. T. Poole and H. Rosenthal. The polarization of american politics. *The journal of politics*, 46(4):1061–1079, 1984.

K. T. Poole and H. Rosenthal. A spatial model for legislative roll call analysis. *American journal of political science*, pages 357–384, 1985.

K. T. Poole and H. Rosenthal. Patterns of congressional voting. *American journal of political science*, pages 228–278, 1991.

K. T. Poole and H. Rosenthal. *Congress: A political-economic history of roll call voting*. Oxford University Press, USA, 2000.

K. T. Poole and H. L. Rosenthal. *Ideology and congress*, volume 1. Transaction Publishers, 2011.

K. M. Quinn, A. D. Martin, and A. B. Whitford. Voter choice in multi-party democracies: A test of competing theories and models. 43(4):1231–1247, 1999. ISSN 0092-5853. doi: 10.2307/2991825. URL https://www.jstor.org/stable/2991825. Publisher: [Midwest Political Science Association, Wiley].

N. Schofield, A. D. Martin, K. M. Quinn, and A. B. Whitford. Multiparty electoral competition in the netherlands and germany: A model based on multinomial probit. 97 (3):257–293, 1998. ISSN 0048-5829. URL https://www.jstor.org/stable/30024432. Publisher: Springer.

B. Shor and N. McCarty. The ideological mapping of american legislatures, 2011. URL https://papers.ssrn.com/abstract=1676863.

A. Smithies. Optimum location in spatial competition. *Journal of Political Economy*, 49 (3):423–439, 1941.

D. Stiers. Spatial and valence models of voting: The effects of the political context. 80:102549, 2022. ISSN 0261-3794. doi: 10.1016/j.electstud.2022.102549. URL https://www.sciencedirect.com/science/article/pii/S0261379422001056.

C. L. Struthers, C. Hare, and R. Bakker. Bridging the pond: measuring policy positions in the united states and europe. 8(4):677–691, 2020. ISSN 2049-8470, 2049-8489. doi: 10.1017/psrm.2019.22.

P. W. Thurner. The empirical application of the spatial theory of voting in multiparty systems with random utility models. *Electoral Studies*, 19(4):493–517, 2000.

P. W. Thurner and A. Eymann. Policy-specific alienation and indifference in the calculus of voting: A simultaneous model of party choice and abstention. 102(1):51–77, 2000. ISSN 0048-5829. URL https://www.jstor.org/stable/30026136. Publisher: Springer.

K. E. Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

Z. Wang, S. Hale, D. I. Adelani, P. Grabowicz, T. Hartman, F. Flöck, and D. Jurgens. Demographic inference and representative population estimates from multilingual social media data. In *The World Wide Web Conference*, WWW '19, page 2056–2067, New York, NY, USA, 2019a. Association for Computing Machinery. ISBN 9781450366748. doi: 10.1145/3308558.3313684. URL https://doi.org/10.1145/3308558.3313684.

Z. Wang, S. Hale, D. I. Adelani, P. Grabowicz, T. Hartman, F. Flöck, and D. Jurgens. Demographic inference and representative population estimates from multilingual social media data. In *The world wide web conference*, pages 2056–2067, 2019b.

J. Wheatley and F. Mendez. Reconceptualizing dimensions of political competition in europe: A demand-side approach. *British Journal of Political Science*, 51(1):40–59, 2021.