MORE VOCAL FOR LOCAL? THE ROLE OF CANDIDATES' PLACE OF RESIDENCE IN THE DUTCH PROVINCIAL AND WATER BOARD ELECTIONS

Doris L. Bukman
6607209

Source: Gemeente Weert

MSc Applied Data Science, Utrecht University

First supervisor: Prof dr. ir. Vincent Buskens
Second supervisor: Dr. Mahdi Shafiee Kamalabad

July 2nd 2023

ABSTRACT

Voting behavior is partly determined by candidate characteristics, one of which is place of residence. The *friends-and-neighbors voting* hypothesis posits that contenders for office garner disproportionally more support from voters in their hometowns. Voters are more likely to know of or know the candidate personally, they may believe that native candidates will defend their interests better, or they become mobilized through a sense of shared identity. While the bulk of the literature is focused on national elections in majoritarian systems such as the US or UK, this thesis project will investigate the role of the hometown advantage for the Dutch Provincial (*Provinciale Staten*) and Water Board (*Waterschaps*) elections. Specifically, it will research *whether candidates receive more votes in the voting stations located in their places of residence in either election*. These elections are second-order, conducted on the Provincial level, meaning that voters have a lower interest in and awareness of them than in the national elections. As voters seek out less information about candidates' merits and stances, they may rely more heavily on characteristics such as background to make political decisions. The Dutch elections are especially suitable for such research, as candidates' place of residence is reported on the ballot, meaning that this information is easily available.

Because of the excess presence of zero-values, the analysis is split into a two-part model to investigate firstly what determines whether or not a candidate receives any votes in either election and secondly how the number of votes is influenced. I find that candidates do receive significantly more votes in voting stations located in their hometowns and that this effect is quite large and consistent between the models. Candidates who are the only ones from a certain town running on their party's ballot receive another vote boost. Furthermore, it appears that this effect is slightly stronger in more rural areas, where community ties tend to be stronger. The roles of wealth and candidate list position on the same-town residence effect remain largely unclear. Similarly, there are no clear convincing differences in the main predictor between either election. Overall, this research offers support for the friends-and-neighbors voting hypothesis in the context of the Dutch Provincial and Water Board elections.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The historic victory of the *BoerBurgerBeweging* (Farmer-Citizen Movement) in the Provincial elections in March 2023 marked an important shift in Dutch political life. The newcomer party (BBB for short) became the biggest in all 12 Dutch provinces (Schelde & Kanne, 2023). This triumph can be understood in light of the recent *nitrogen crisis*: to cut nitrogen emissions in half by 2030, the Dutch government needs to introduce drastic measures that target agriculture, the main culprit (Halm, 2022). The BBB was founded in support of the widespread farmers' protests against these policy decisions in 2019 (*Partijgeschiedenis*, 2021). However, it has grown to represent much more. Support for the BBB is seen as an expression of the electorate's general discontent with mainstream politics, specifically the sentiment that rural interests have been neglected by the national government (Schelde & Kanne, 2023). This is evident in the stark split in the success of the BBB between rural and urban areas (Klaassen, 2023). Voters believe that the rural identity is underrepresented, and the BBB 'really understands what resonates with farmers' (Klaassen, 2023).

The factors that determine political behavior are complex and context-dependent. Citizens vote not only to further their rational self-interest but also to reinforce their identities (Ansolabehere & Puy, 2016). Group membership influences how voters perceive the consequences and the morality of policy decisions (Conover, 1984). Examples of such characteristics are religion, language or gender. The group membership of politicians may serve as a heuristic for making electoral decisions, as voters rarely have complete knowledge of their own and different parties' positions on relevant topics (de Leeuw, 2017; Caprara & Zimbardo, 2004). Some scholars argue that there is an increased personalization of politics in Western democracies in the last few decades: there is a stronger focus on politicians and their traits, rather than on parties, issues, and ideologies (Adam & Maier, 2010). There is a distinction to be made between first and second-order personalization. The former concerns party choice, while the latter refers to the choice for a certain candidate within a certain party, as is the case in proportional representation (PR) systems (Van Holsteyn & Andeweg, 2012).

Because voters have already made an ideological choice during the stage of second-order personalization, it is an interesting avenue to study the congruence between voters' and candidates' traits. One such trait is geographical proximity or locality. If voters consider their place of residence to be a salient part of their identity, it may influence their likelihood to vote for a candidate with a similar background. This geographical identity can be very broad, such as the rural-urban division surrounding the Dutch 2023 Provincial elections. The public discourse about the nitrogen crisis activates this part of citizens' identity, and prompts them to care more about local issues. Another example of this mechanism is how the economic crisis incentivized Irish voters to prioritize national issues over local issues (Put et al., 2020). This sense of identity can also be tied to the town-of-residence specifically, with citizens finding it desirable to have their hometown represented in the electorate. Such dynamics differ geographically, and may be bigger in rural areas or certain cities for instance, or vary along with factors such as education and income.

This thesis project will investigate the role of candidates' locality in Dutch politics. Specifically, it will look at whether candidates' town-of-residence influences the number of votes cast they receive in those towns, for two concurrent Dutch elections: the 2023 Provincial (*Provinciale Staten*) and Water Board (*Waterschappen*) elections. The choice for two subnational elections allows for a comparative investigation of locality for elections about which voters have generally less knowledge. I will analyze this phenomenon using various variables such as the degree of urbanization in a certain town, party list composition, and wealth. I begin with a short description of the Provincial and Water Board electoral systems. I then continue with a review of the literature in the theoretical framework, from which I formulate my final research question and hypotheses. In the methodology section, I detail the data collection steps and the statistical analysis, which consists of separating the data into a two-part model. Afterward, I report the results and examine their implications in the discussion section.

2. BACKGROUND: PROVINCIALE STATEN AND WATERSCHAPPEN

The Dutch political system is divided into three levels of governance, with the national parliament at the top, and the municipalities at the bottom. In between we find the Provincial level, which consists of two electoral bodies: the Provinciale Staten and Waterschappen. This threefold system has an hourglass shape, as the middle layer has relatively little political power (Lefevere & Van Aelst, 2014). Elections for both occur every four years and since 2015 they occur on the same day to improve turnout (Havekes & Elfferich-Rodenburg, 2020). Dutch voters generally have little knowledge of and interest in the Provincial (PS) and Water Board (WB) elections (Vollaard et al., 2021). Furthermore, turnout for both is relatively low, just over 50% (Vollaard et al., 2021). While local parties are running in both elections, the majority of the votes go to national parties, especially in the PS elections (Binnema & Vollaard, 2021; Vollaard et al., 2021). They are based on open-list proportional representation (PR), as is the case for every Dutch election (Binnema & Vollaard, 2021). Allocation of candidates to seats is based on the list order. However, those placed lower on the list can pass higher candidates if they receive more votes and the number of votes surpasses 25% of the *kiesdeler*: the total votes cast in the election divided by the number of seats (Van Holsteyn & Andeweg, 2012).

The Netherlands is divided into 12 different provinces, each of which has its own parliament called the Provinciale Staten. This legislative body determines policies and oversees their execution carried out by the *gedeputeerde staten* (*Provinciale Staten*, n.d.). It also chooses the members of the national senate. The Provinciale Staten is concerned with topics such as spatial planning, housing, and air quality (Koninkrijksrelaties, 2016). While there are 12 provinces, some of these are divided into different electoral districts called *kieskringen,* 20 in total. Political parties can provide a different ballot of candidates for each kieskring, or choose to only run in certain kieskringen. The votes are aggregated per province (Kiesraad, 2016b).

The Water Boards are tasked with water management, such as regulating dykes or water purification (Kiesraad, 2016a). Part of its members is elected, and a part of the seats is reserved for representatives from agricultural and environmental groups (Kiesraad, 2016a). There are 21 Water Boards in total, the borders of which vary substantially from the provinces. As a result, certain municipalities and towns are part of multiple Water Boards. For instance, the municipality of Rotterdam is divided into *Waterschap Delfland*, *Schieland en de Krimpenerwaard,* and *Hollandse Delta*.

3. THEORETICAL FRAMEWORK

In their study of US Southern politics, Key and Heard (1949) remarked that candidates running for Governor pull large victories in their home areas. Put bluntly: 'he [the candidate] gains support, not primarily for what he stands for or because of his capacities, but because of where he lives' (Key & Heard, 1949, 37). This observation prompted a tradition of research into the effects of so-called *friends-and-neighbors* voting: candidates appear to garner disproportionally more support closer to home (Put, 2021). Three possible mechanisms explain this effect. Firstly, voters in a certain area have more access to information about a candidate that lives close to them (Górecki et al., 2022). Citizens may be more aware of the existence of the candidate or may know them personally. This differential access decreases as distance increases. Secondly, voters may hold implicit beliefs about the 'behavioral localism' of candidates, that native politicians may defend the interests of the area or city better (Put, 2021). A candidate's place of residence may be an informational shortcut about the kind of leadership that can be expected. Thirdly, voters may choose candidates from their hometown due to the sense of a shared identity. Certain groups feel more affection towards politicians based on their objective characteristics, such as their gender or their place of residence (Plutzer & Zipp, 1996). It may be beneficial for politicians to emphasize this part of their identity to garner more votes.

Research into this spatial-political phenomenon is very well-established for majoritarian systems like the US and UK, and recently there is also more attention paid to list PR systems (Put et al., 2020). Especially in candidate-centered electoral systems, voters rely on 'personal vote-earning attributes' as cues to make political decisions, as there is more choice than in majoritarian systems (Put et al., 2019). This allows for *intra-party competition*, as candidates try to get elected through preference votes. This mechanism is stronger in open list systems than in closed list systems, as candidates are competing individually for a seat, rather than as a party (Shugart et al., 2005). If a friends-and-neighbors mechanism is present, candidates on a list for a particular party can expect more votes from party supporters from their region of residence. If the mechanism is strong enough, it could even sway voters from one political party to another. Effects of local preference have been found for elections in Norway (Fiva & Halse, 2016), Finland (Put et al., 2020), Lithuania (Herron & Lynch, 2019), and Ireland (Górecki & Marsh, 2012) among others. However, there is little research on localism mechanisms in Dutch elections. One exception is Van Holsteyn and Andeweg (2012), who concluded that in the national elections of 2010, region did not seem to have a large influence.

Furthermore, while the bulk of the literature is focused on national elections, authors such as (Górecki et al., 2022) turn to subnational localism mechanisms. Local elections are often considered to be 'second-order elections' meaning that they are seen as less important than 'first-order elections,' typically on the national level (de Leeuw, 2017). Second-order elections are characterized by different voting behavior, such as lower turnout and a preference for smaller parties (Marien et al., 2015). Simultaneously, second-order elections may be defined by first-order considerations. For instance, voters often use a local election to 'punish' parties in government, as is the case for the Dutch 2023 Provincial elections (Marien et al., 2015). However, not all local issues can be interpreted through the lens of national politics and voters may feel stronger ties to candidates from the same place of residence during local elections. Additionally, as voters seek out less information during second-order elections, place of residence may be a more potent shortcut to making electoral decisions (Lefevere & Van Aelst, 2014).

This research thus contributes to the literature by firstly focusing on the Netherlands, an electoral system in which friends-and-neighbors voting has yet to be thoroughly explored. Secondly, it investigates second-order rather than first-order elections. Specifically, it will take into account two subnational elections: the Provincial (*Provinciale Staten*) and Water Board (W*aterschaps*) elections. These are most likely cases for the existence of a friends-and-neighbors mechanism. Since the elections are more local than the national elections, we can expect voters to feel stronger ties to local candidates. This is somewhat supported by the finding by Hessing (1985) that more preference votes are cast for the Provincial than the national elections. This difference could be due to stronger localism effects. Furthermore, because interest in and awareness of these elections is lower than the national or municipal elections, we can expect voters to use place-of-residence as an informational shortcut as the basis for the vote, rather than extensive information about a candidate's qualifications and positions. This is especially easy for Dutch voters, as the ballots report each candidate's place of residence next to their name.

Moreover, comparing two concurrent elections allows for a deeper exploration of the meaning of locality for voting behavior. The Provincial and Water Board elections differ in three potentially important ways. Firstly, there is more awareness of the Provincial elections (Vollaard et al., 2021). The Water Board elections have a slightly lower turnout, but this difference was more dramatic before 2014. From then on, the elections were made concurrent and both ballots could be cast at the same time (Vollaard et al., 2021). Importantly, because the Provincial States elect the members of the Dutch Senate, they have a direct impact on national politics (Binnema & Vollaard, 2021). Secondly, both elections are conducted in different geographical units. The positioning of the Water Boards is drastically different than the Provinces, with little regard for administrative boundaries. This could lead to a lower sense of geographical belonging than for the Provincial elections. Thirdly, the duties that each administrative body is tasked with may incentivize different voting behaviors. The Water Boards deal mainly with hydrologic management, such as supplying clean water and keeping dykes safe (Koninkrijksrelaties, 2011). The Provincial States, on the other hand, deal with a broader set of tasks, although their influence is small (Binnema & Vollaard, 2021). Therefore, voters may consider a candidate's background to be more important in one election than the other.

The effect of candidates' place of residence is not uniform, and is possibly contingent on factors such as the degree of urbanization. Several authors find that local effects are stronger in rural or suburban communities (Arzheimer & Evans, 2012; Put, 2021; Put et al., 2020). This is due to the stronger sense of community in smaller, rural areas (Put, 2021). Urban voters are also less likely to stay in the same place and thus have a weaker sense of local identity (Put et al., 2020). Additionally, the effect may be regulated by demographic factors like education or wealth. Political sophistication is considered negatively related with characteristic-based voting (Nagtzaam, 2019). In areas where residents have more interest and resources to invest in politics, such as wealthier areas, the effect of same town residence may be lower. Furthermore, candidates could also expect the vote boost from the friends-and-neighbors effect to diminish as there are other candidates from the same area on the list (Van Erkel, 2019). In strongly localized electorates such as Ireland, political parties take this into account by spreading candidates over many constituencies to maximize the total amount of votes the party gets (Kavanagh et al., 2021). Additionally, we can expect local effects to differ based on a candidate's position on the list. Candidates lower on the ballot are typically less well-known, and a larger portion of their votes may be due to their background. Some authors control for or exclude a candidate's position as party leader, as they receive substantially more votes (Van Erkel, 2019). According to Put's (2021) multi-country analysis, there is a friends-and-neighbors effect for party leaders as well, albeit with a different dynamic than for other candidates. In the Dutch system, we can expect the candidates above the kiesdeler to receive more votes, since they have a higher chance of ending up in office. Similarly, it is possible that the candidates just below this line also receive more preference votes. Lastly, the last candidate on the list is traditionally a well-known person who boosts the party's popularity, but who is not actually interested in getting elected (*lijstduwer*).

Many authors measure locality in terms of distance between the candidate and the voters (Arzheimer & Evans, 2012; Gimpel et al., 2008; Put et al., 2020). This is especially useful when researching the effects of personal ties that candidates have with their neighbors, or the knowledge that voters have of the candidates running close to them. However, such address data is not available for this project, and I instead use the name of the current place-of-residence of the candidate. Therefore, friends-and-neighbors voting is measured more in terms of the sense of identification that voters feel with a candidate with a similar background, and their beliefs about whether these candidates will represent the interest of the town better. Where a candidate currently lives also tells us very little about where a candidate grew up, lived for the majority of their political career, or their ties to the town. Fortunately, it is very easy for voters to identify the location of candidates, as their town-of-residence is included on the ballot. This allows them to promote local candidates without prior knowledge. It is also still more likely that the voter knows the candidate, if they live in the same town. Furthermore, the use of station-based election results, rather than survey data, to make inferences about friends-and-neighbors risks the ecological fallacy: relationships for the aggregate level (e.g. the voting station) may not exist on the individual level. This limitation must be kept in mind when analyzing the results.

3.1. Hypotheses

I am primarily interested in whether *candidates in the PS and WB elections receive more votes in their hometowns.* The literature review above yields the following set of hypotheses:

> H1: Candidates receive more votes in a certain voting station if they are residents of the place the station is in.
>
> H2: Candidates receive more votes in a certain voting station if they are only residents of the place the station is in running on their party's ballot.
>
> H3: The effect of same-town residence is stronger in the less well-known Water Board election.
>
> H4: The effect of same-town residence is stronger in rural areas.
>
> H5: The effect of same-town residence is weaker in wealthier areas.
>
> H6: The effect of same-town residence is stronger for candidates lower on the ballot.

4. METHODOLOGY

4.1. Data collection

The *Kiesraad* collects and publishes the ballot counts per voting station for each Dutch election (Kiesraad, 2023). The data for both elections are separated into three different datasets, two of which I used: a breakdown of the ballot counts per candidate for each polling station and information about each candidate running in each Kieskring or Water Board. Combined with data on the locations of the voting stations, and data about the demographic characteristics per four-digit postal code, I merged these into two final datasets for each election respectively. Each row represents the votes for a candidate in a certain voting station, along with the candidate's characteristics such as their place-of-residence, and characteristics of the area around the voting station. I removed the party's leaders (*lijsttrekkers*) from the analysis altogether, as they could distort the analysis. A detailed description of the data sources and the data handling steps I undertook can be found in Appendices 1 and 2.

The structure of the original dataset posed two problems. Firstly, the municipalities individually report the results to this agency, and there are no global quality checks done on the data (Kiesraad, 2023). This possibly explains the missing data issue in the ballot counts dataset: for around 760 voting stations the postal code is left out. This is problematic because while it is still possible to deduce the station's Kieskring or Water Board, it is unknown where the station is. It renders around 8% of the data unusable for the analysis, in both datasets. Because it appears as though this missingness is not clustered geographically, I assume that it does not substantially threaten the validity of the results. Secondly, while for the Provincial elections each voting station is only part of one kieskring, one voting station can receive ballots for multiple Water Boards as some municipalities are part of two or three Water Boards.[1] These stations thus collect votes for multiple Water Boards, and it is more difficult to deduce which votes are for which since this is not recorded in the dataset. To solve this issue I assigned the voting stations based on the composition of the ballot, rather than the postal code.

4.2. Statistical design

The distribution of the data is unsuitable for standard regression techniques, due to two reasons. Firstly, there is an excessive presence of zeros in the ballot counts. In an average voting station, the majority of the running candidates do not receive any votes, due to the generally low turnout and multiplicity of choice (see Figure 1). This disbalance is more pronounced in the Provincial elections, for which there

---

[1] In the majority of these municipalities the voting stations are divided into the different Water Boards, but in a few places, such as Amstelveen, voters can cast their ballot in every station in the municipality, no matter the Water Board that they live in.

are more running candidates. As a result, the data is strongly right-skewed data, which violates key assumptions such as normality and homoscedasticity. Various statistical techniques can be used to deal with this, and which is more appropriate is partly determined by the process that generates the zero values. We could be dealing with so-called 'floor effects' where the concentration of zeroes is explained by a faulty metric that captures most variability on the lower end of the scale (von Klipstein et al., 2023). In this case, the Tobit regression model is recommended, in which the zero values are assumed to be related to a latent variable, the probability of which is modeled (Boulton & Williford, 2018). On the other hand, if zeros represent 'true zeros,' as is the case in the election data, a two-part model is preferred. Here, the regression analysis is split into two components: a logistic regression for a variable that represents whether or not an observation is zero, and a linear regression for only the nonzero values (Boulton & Williford, 2018). The underlying assumption is that the process which generated the zero values is different from the process that generated the nonzero values. There are more sophisticated options, such as zero-inflated negative binomial or Poisson regression, but these are only suitable for count data, not for proportions which are used in this analysis to remove the effect of voting station size (Liu et al., 2019).

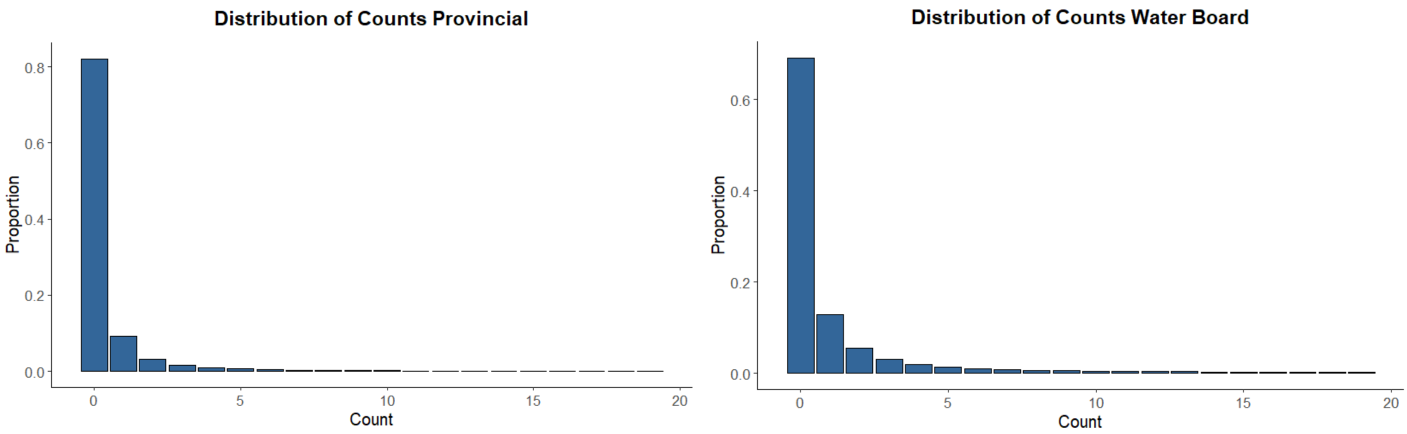

Figure 1. Distribution of vote counts under 20 for Provincial and Water Board elections

Secondly, once we separate the zero values, we are still left with a strongly right-skewed distribution. To address this problem, it is conventional practice to use operations such as a square root or log transformations (Geissinger et al., 2022). Performing these transformations on the count data yields improvements, but does not sufficiently deal with the skew, and the coefficients cannot be interpreted easily in terms of the pre-transformation data (Ferrari & Cribari-Neto, 2004). Therefore, I employ a Beta regression, which handles asymmetry well and is fit for proportion data (Chai et al., 2018). This approach utilizes a beta probability density function, that is flexible to a multitude of shapes (Geissinger et al., 2022). I will combine the two approaches, by first conducting a logistic regression for the zero and nonzero values, and then a separate Beta regression for the nonzero values. This results in four models: logistic regression for the Provincial elections (logPS), beta regression for nonzero values for the Provincial elections (betPS), and the same for the Water Board elections (logWB and betWB). Because of possible clustering, I will include random effects for the municipality and candidate names, which can be integrated using the *glmmTMB* package in R.

By assessing the impact of the fixed effect variables on the random effects and the overall model performance, I will decide upon the model fit. After this process, I will rerun the analysis using the final specification, for each of the 10 largest political parties, and each of the kieskringen and water boards as a consistency check. Another hurdle is the strategy for testing the significance of mixed models. R packages *lme4* and *glmmTMB* usually do not output p-values, because of the difficulty of deciding upon degrees of freedom with random effects (Luke, 2017). Two commonly used methods for estimating significance in mixed models are Wald and likelihood ratio tests (LRT), the latter being preferred (University of Wisconsin-Madison, n.d.). However, because conducting LRT for each variable in each model, amounting to 64 tests in total, is computationally expensive and time consuming for datasets of this size, I rely mainly on the t-value outputs of the regression analyses. This approach applies the z distribution to the Wald t-values (t-as-z). While this method is less accurate for smaller sample sizes, it yields similar results as other methods with large datasets (Newsom, 2019). After comparing the LRT and Wald results for one of the regression models, I found no differences in the reported significance levels. For the sake of accuracy, I conduct LRT tests for variables whose t-values are close to the cusp.

4.3. Operationalizations

The variable operationalizations can be found in Table 1 below.

Table 1. Variable descriptions

Name	**Description**	**Format**
Dependent variables		
Count logistic models	Whether a candidate received zero, or more than zero votes	Binary variable [0, 1]
Count beta models	For candidates that received more than zero votes in a certain voting station, the proportion of their respective party's total votes.	Continuous variable scaled <0, 1>
Independent variables		
Place of residence[2]	Whether a candidate is a resident in the location of the given voting station. This location is a town or city that has been denoted as such by the municipality it is part of.	Binary variable [0, 1]

[2] Place of residence is also referred to as, 'same-town residence,' 'locality' or 'hometown effect' throughout this thesis.

Only candidate	Whether a candidate is the only candidate on their party's list from the given town or city.	Binary variable [0, 1]
Urbanization	Degree of urbanization in a given postal code, measured as the number of addresses per km2. The least urban postal codes have fewer than 500 addresses, and most urban 2500 or more.	5-level ordinal scale
Wealth	Average property worth (*woz waarde*) in a given postal code in 100,000€. This is an imperfect measure, as it does not consider variability within the area or other aspects of wealth such as income.	Discrete variable ranging from 0.82 to 14,48
Lijstduwer	While party leaders have been excluded from the analysis altogether, this variable denotes whether a candidate is the last candidate on the party's list.	Binary variable [0, 1]
List position	Denotes a candidate's position on the party list as either high (top 10%), middle, or low (bottom 10%). In reality, we would expect these categories to depend on the expected number of seats a party receives, but this polling data is not easily available, and the percentages are instead bluntly estimated.	3-level categorical variable

5. RESULTS

The results are divided into the estimates for the four different models. Each is reported in three versions: a model with only the random effects (0), a model where the predictors have been added (1), and a model with the interaction effects as well (2). The data structures meet the assumptions of logistic and beta regression. More details about the model diagnostics can be found in Appendix 3.

5.1. Logistic regression for Provincial elections (logPS)

Table 2 presents the outputs of the three regression variations of the zero values logistic regression. For each variable, the estimated log odds, the level of significance derived from conducting the LRT, and the standard error are reported. The variance explained by municipality random effects was negligible, and this has therefore been excluded from the model. LogPS1 and 2 show that place of residence has a significant and large effect on the log odds of receiving any votes. Compared with the 'baseline candidate' who is not from the town that the voting station is in, the predicted probability of receiving any votes increases from 45,8% to 95,5% in logPS1.[3] Furthermore, candidates who are the only ones running for their party from a certain place receive another significant boost in their chances. The covariate urbanization has no significant effect when included in model 1. The effect of wealth on log odds is very small and negative, but significant. The covariates that denote list position have a clear effect on the log odds of receiving a vote, again as expected.

The role of these variables becomes more interesting when we include interaction effects with our main predictor in logPS2. Here, there is a statistically significant effect of the degree of urbanization. As was hypothesized, an increase in urbanization is associated with a (small) decrease in the strength of the relationship between place of residence and voting. Interestingly, there is a similarly small but positive effect of wealth on the strength of the relationship. As was expected, being the last candidate on the list (lijstduwer) leads to a significant increase in the log odds of receiving more than zero votes. Notably, it appears that the effect of same-town residence becomes stronger near the middle of the party's list. There is no significant difference for candidates at the bottom of the list. Adding the predictors leads to a lower AIC (Akaike Information Criterion), indicating a better fit. While the interaction effects are significant, adding them does not change the goodness-of-fit.

Table 2. Estimates for logistic regression of zero values PS elections

Name	**logPS0: RE only**	**logPS1: all variables**	**logPS2: interactions**
Intercept	-1.85*** (0.017)	-0.073 . (0.038)	0.065. (0.038)
Place of residence	-	3.229*** (0.034)	2.873*** (0.119)
Only candidate	-	2.201***	2.185***

[3] The baseline candidate is at the top of the list, and is from a town with mean wealth and mean urbanization.

		(0.068)	(0.068)
Urbanization	-	≈0.000 (0.001)	0.001 (0.001)
Wealth	-	-0.030*** (≈0.000)	-0.032*** (≈0.000)
Lijstduwer	-	0.727*** (0.053)	0.718*** (0.053)
List position			
Low position	-	-2.460*** (0.059)	-2.443*** (0.059)
Middle position	-	-2.094*** (0.038)	-2.100*** (0.038)
Urbanization * Place of residence	-	-	-0.040*** (0.009)
Wealth * Place of residence	-	-	0.023*** (≈0.000)
List position * Place of residence			
Low position	-	-	-0.132 (0.150)
Middle position	-	-	0.473*** (0.121)
Candidate random intercept	1.861	1.561	1.561
AIC	3824641	3171074	3171074

Level of significance:

.	0.10
*	0.05
**	0.01
***	0

5.2. Beta regression for the Provincial elections (betPS)

I now turn to the results of the analysis for non-zero vote count proportions, which can be found in Table 3. Similar to the logPS model, place of residence exhibits a strong and significant effect on the number of votes. Being the only candidate from a town also has a significant impact. There are also notable differences with the logistic model. Wealth no longer has a significant effect on the vote count, either as a single variable or in the interaction with place of residence. Urbanization, on the other hand, does appear to have a small but significant impact. This is the inverse of the logPS model. The interaction effects for list position show another difference: the effect of same-town residence decreases significantly as candidates move down the list, in the middle as well as the low category. The AIC values of the three models show that there is a small improvement in goodness-of-fit after adding the interaction effects.

Table 3. Estimates for beta regression of nonzero values PS elections

Name	betPS0: RE only	betPS1: all variables	betPS2: interactions
Intercept	-4.813*** (0.004)	-4.924*** (0.010)	-4.899*** (0.010)
Place of residence	-	0.759*** (0.015)	1.297*** (0.004)
Only candidate	-	0.859*** (0.022)	0.906*** (0.023)
Urbanization	-	0.042*** (0.001)	0.045*** (0.001)
Wealth	-	-0.001 (0.001)	-0.001 (0.001)
Lijstduwer	-	0.064*** (0.018)	0.068*** (0.018)
List position			
Low position	-	-0.359*** (0.017)	-0.402*** (0.017)
Middle position	-	-0.304*** (0.010)	-0.341*** (0.010)
Urbanization * Place of residence	-	-	-0.041*** (0.002)
Wealth * Place of residence	-	-	0.003 (0.002)
List position * Place of residence			
Low position	-	-	-0.646*** (0.054)
Middle position	-	-	-0.475*** (0.039)
Candidate random effects			
Random intercept	0.061	0.055	0.054
AIC	-6353442	-6628056	-6628515

Level of significance:
. 0.10
* 0.05
** 0.01
*** 0

5.3. Logistic regression for the Water Board elections (logWB)

Next, I turn to the results of the same set of regression equations for the Water Board elections. Upon first inspection, the results are very similar to the Provincial data. The estimates of same-town residence are of comparable size. In this model, the covariates wealth and urbanization are both significant and negative, even though they are small. Notably, there is no longer a significant interaction effect between urbanization and the main predictor, although the reason for this deviation is unclear. Similarly, the interaction between wealth and same-town residence is much less strong. Interestingly, the effects of being the last candidate on the ballot and of the list position variables are substantially bigger than for the Provincial elections. The results of list position in combination with the main predictor are similar to the results of logPS. Candidates on the bottom of the list see a non-significant increase in the effect of same-town residence, while the middle category candidates receive significantly more benefits in their home towns. Overall, the model variations are very comparable. The variance explained by random effects is only slightly smaller. The goodness-of-fit changes between the three models follow a similar pattern, but there is a small difference in AIC values between logWB1 and 2, while they remained the same for the Provincial elections.

Table 4. Estimates for logistic regression of zero values WB elections

Name	logWB0: RE only	logWB1: all variables	logWB2: interactions
Intercept	1.354*** (0.047)	1.354*** (0.048)	1.362*** (0.047)
Place of residence	-	3.332*** (0.045)	3.037 *** (0.162)
Only candidate	-	1.628*** (0.088)	1.623 *** (0.088)
Urbanization	-	-0.006*** (0.002)	-0.005** (0.002)
Wealth	-	-0.025*** (0.002)	-0.026*** (0.002)
Lijstduwer	-	1.919*** (0.093)	1.926*** (0.093)
List position			
Low position	-	-3.874*** (0.080)	-3.867*** (0.080)
Middle position	-	-2.614*** (0.048)	-2.621*** (0.048)
Urbanization * Place of residence	-	-	-0.013 (0.012)
Wealth * Place of residence	-	-	0.0012 . (≈0.000)
List position * Place of residence			
Low position	-	-	-0.213 (0.198)
Middle position	-	-	0.371 * (0.164)
Random intercept	1.611	1.434	1.434
AIC	2116460	1783524.6	1783506.4

Level of significance:
. 0.10
* 0.05
** 0.01
*** 0

5.4. Beta regression for the Water Board elections (betWB)

The results for beta regression for the Water Board elections are again similar to their Provincial counterpart. Remarkably, the estimated coefficient of place of residence is considerably larger (1.091 versus 0.759). Given the standard errors, this difference is significant. Similarly, lijstduwer and list position have stronger negative effects, just as in the logistic regression. There is again a more consistent effect of list position on same-town residence compared to logPS: candidates at the bottom and middle of the list receive relatively fewer votes in their hometowns compared to their peers at the top.

Table 5. Estimates for beta regression of nonzero values WB elections

Name	**betWB0: RE only**	**betWB1: all variables**	**betWB2: interactions**
Intercept	-4.395***	-4.247*** (0.015)	-4.241*** (0.015)
Place of residence	-	1.091*** (0.020)	1.569*** (0.049)
Only candidate	-	0.965*** (0.028)	0.970*** (0.028)
Urbanization	-	0.021*** (0.001)	0.022*** (0.001)
Wealth	-	0.005*** (0.001)	0.009*** (0.001)
Lijstduwer	-	0.157*** (0.032)	0.149*** (0.032)
List position			
Low position	-	-0.711*** (0.029)	-0.744*** (0.029)
Middle position	-	-0.601*** (0.015)	-0.621*** (0.015)
Urbanization * Place of residence	-	-	-0.016*** (0.003)
Wealth * Place of residence	-	-	-0.010*** (0.002)
List position * Place of residence			
Low position	-	-	-0.739*** (0.067)
Middle position	-	-	-0.411*** (0.048)
Candidate random effects			
Random intercept	0.085	0.089	0.090
AIC	-3939851	-4151788	-4151958

Level of significance:
. 0.10
* 0.05
** 0.01
*** 0

5.5. Summary of model results

The results of the four regression models laid out above show a relatively consistent picture: same-town residence does significantly increase the vote count of candidates running in the Provincial and Water Board elections. While this effect is significantly stronger in the Water Board elections in the case of the beta regressions, no such difference is visible for the logistic models. We also see that candidates who are the only ones from a certain place of residence on their party's ballot receive an extra vote boost in their hometowns. Urbanization and wealth of the postal code around the voting stations only very slightly impact vote proportion, and not consistently. This result can be expected in the logistic regressions, as higher urbanization means more voters and, therefore, more candidates that receive votes, for instance. For the beta regressions, however, this effect is not so intuitive, as the dependent variable is the proportion of votes, and is thus not related at all to how many total votes are cast. This could be related to unaccounted interactions between these covariates and other variables, but this has not been verified.

In interaction with same-town residence, urbanization has a small but significant negative effect, which confirms that as towns are more rural, the friends-and-voters effect increases. There is one exception, as no significant effect was found for the zero-value regression for the Water Board elections (logWB). The interaction effect of wealth is small and usually positive, but only significant in the logistic models. Candidates' list position strongly influences the number of votes as a fixed effect. Its interpretation as an interaction with place-of-residence is murkier. In both beta regression models (betPS and betWB), the lower the candidate is placed, the smaller the hometown advantage is. In the logistic models, on the other hand, the interaction effect is positive and significant for candidates in the middle of the list, and negative and not significant for the bottom candidates in both elections. This, along with the wealth interaction effects, indicates that there is a different mechanism that determines whether a candidate receives zero or any votes, versus what determines different outcomes for the candidates that do receive votes.

5.6. Results for different data subsets

To check the consistency of the models, I ran logPS1 and logWB1 for the zero and non-zero data in different subsets of data: first for the biggest political parties in either election, second for each Kieskring and Water Board. The random effects were omitted, as including them exceeded an acceptable execution time. The coefficient sizes for the largest parties running in both elections are displayed in Figure 2. The full results can be found in Appendix 3. The effect of same-town residence remained significant and relatively large in each subset of the data. The estimate was the largest among the political parties with rural voter bases, such as the BBB, SGP, and CDA. This finding again suggests that there is a stronger preference for local candidates in rural areas.

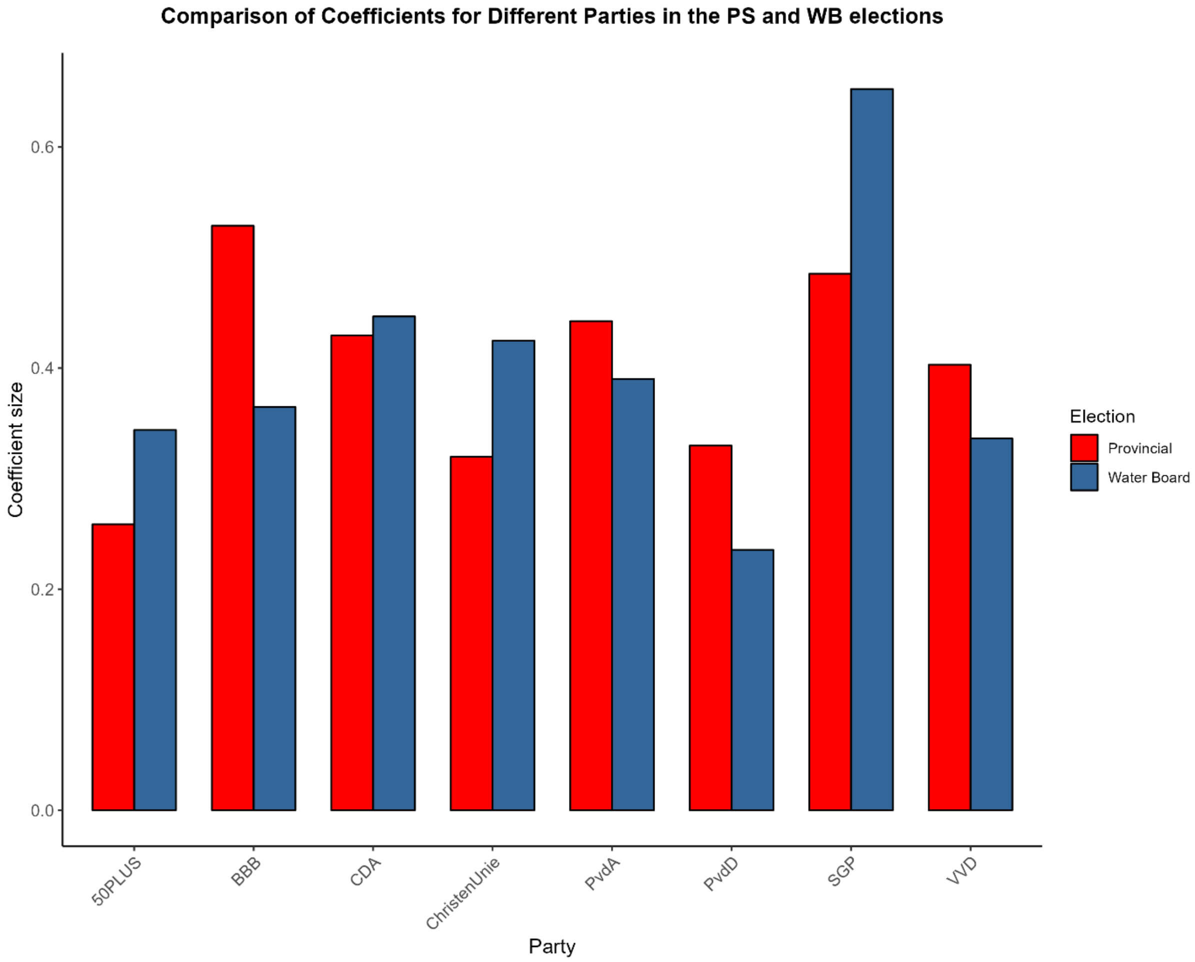

Figure 2. Coefficient sizes same-town residence of logPS and logWB for different political parties

For descriptive purposes, I reran a strongly simplified version of the logistic models for each different municipality, which is mapped in Figures 3 and 4. The results for the beta regression per municipality are similar, and therefore not presented in a map. The grey-striped areas denote the municipalities with missing data due to the absent postal codes for the voting stations, or where there are no native candidates. Notably, the areas with the highest degree of locality voting correspond roughly to the Dutch 'bible belt,' a strip of the Netherlands with a high degree of conservative Reformed Protestant residents. Furthermore, same town residence is also highly influential in the *Waddeneilanden*, the islands in the North of the Netherlands, as well as in other rural areas. These findings are especially true in the Provincial map, while the highest values for the Water Board map are spread more evenly. The major cities such as Amsterdam and Utrecht exhibit far lower levels of locality in both elections.

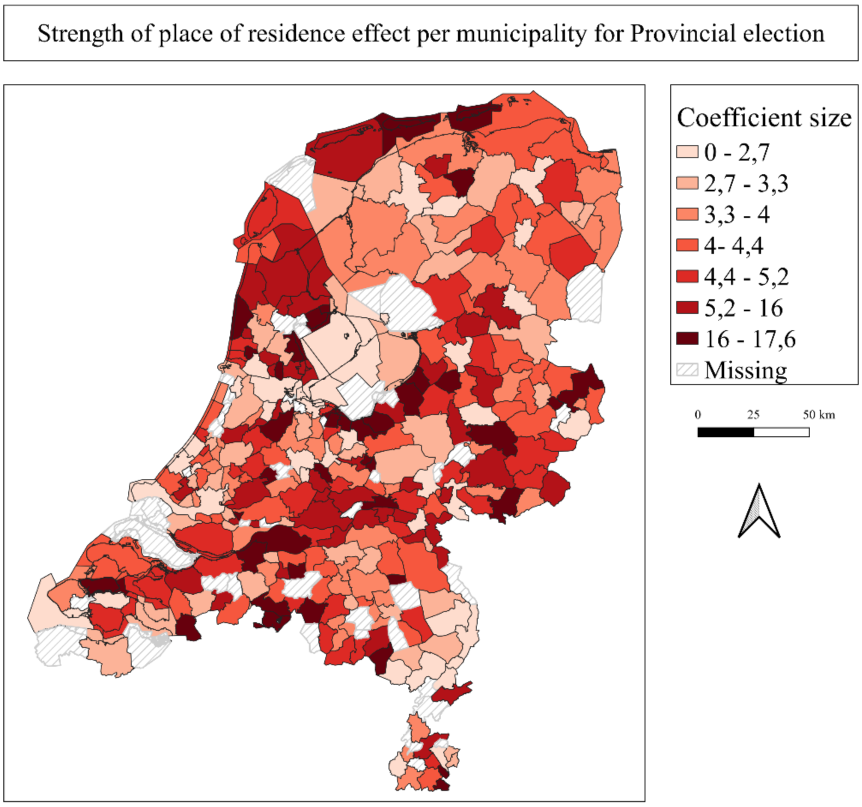

Figure 3. Estimate size of same-town residence on whether or not a candidate receives votes per municipality for the Provincial elections (logPS)

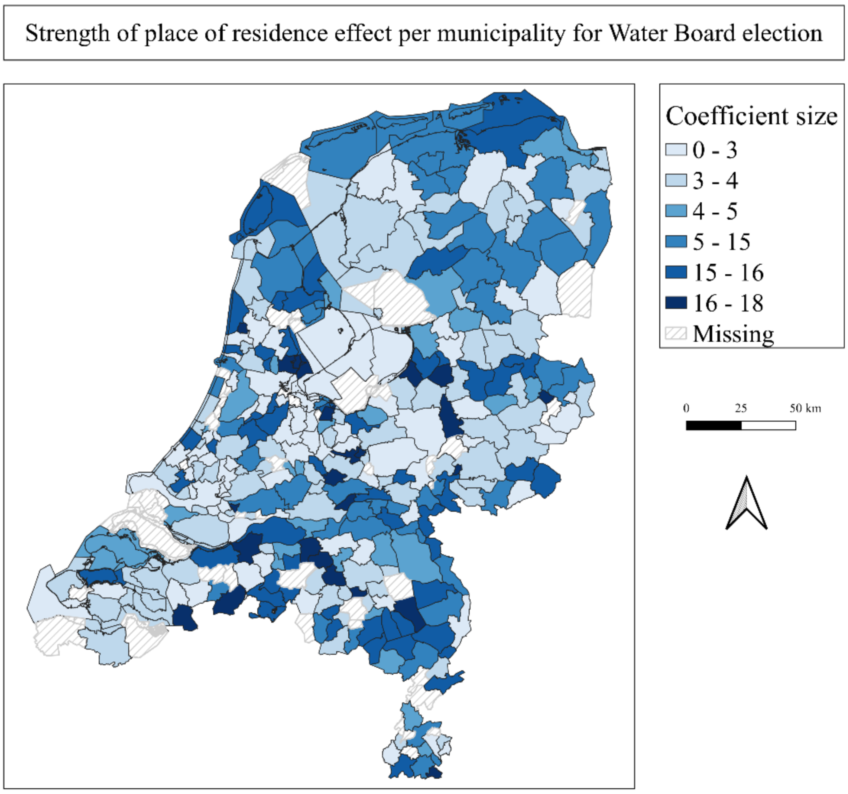

Figure 4. Estimate size of same-town residence on whether or not a candidate receives votes per municipality for the Water Board elections (logWB)

6. CONCLUSION

Overall, the results of the various models indicate that same-town residence does increase vote counts for both the Dutch Provincial and the Water Board elections of 2023. Candidates receive significantly more votes at a specific voting station if they reside in the same place where the station is located (H1). This finding is true in each different Kieskring or Water Board and also holds when separating the data for the largest political parties. Furthermore, candidates receive an extra boost in votes if they are the only candidate from that location running on their party's ballot (H2). The effect of same-town residency appears to be larger for the nonzero values in the Water Board elections (betWB), but not for the logistic regression (logWB) (H3). Therefore, it cannot be concluded that this effect is more important than in the more well-known Provincial elections. One clear finding, on the other hand, is the influence of urbanization on the strength of the residency effect. While estimate sizes are small, there is a significantly lower effect of same town residence as the surrounding postal code becomes more densely populated (H4). These effects are observed again, albeit not statistically tested, in a mapping of the predictor's effect for each municipality. The areas of the Netherlands that are more rural, or where communities are tight, such as in the bible belt or on the Wadden islands show a stronger preference for local candidates. Increased wealth of the voting station area, on the other hand, did not weaken the place of residence effect (H5). Lastly, I found no consistent support for the hypothesis that locality is more important for candidates lower on the party's ballot (H6). In most cases, this effect appeared to be weaker on the middle and bottom of the list. This could be related to the fact that candidates on the top receive more votes overall, which possibly distorts the analysis.

7. DISCUSSION

The results of this research show support for the friends-and-neighbors voting hypothesis. While the bulk of the literature focuses on national elections in majoritarian systems such as the UK and USA, the results confirm that candidates also garner significantly more hometown support in second-order PR elections such as the Dutch Provincial and Water Board elections. The strength of this effect differs across different contexts, but remains salient. What exactly stimulates Dutch voters to prefer local candidates remains largely unexamined. It could be that voters know candidates running in their town personally, especially if community ties are strong. It could also be related to how much voters identify with their place of residence, or to the belief that local candidates will defend the interests of their town better. Regardless of the underlying causal structure, this effect appears to be larger in rural areas, even though this difference is small.

There are several caveats of the research design, which can be divided into concerns related to the data structure, and to statistical choices. Regarding the former, it needs to be noted again that it did not consider the distance between voters and candidates when measuring locality, which would allow for a deeper investigation of the role of candidates' backgrounds. It could for instance be used to measure whether voters also prefer candidates from nearby towns. Secondly, because the data is aggregated on the voting station level, it is impossible to make clear inferences about individuals' voting behavior without risking the ecological fallacy. Thirdly, while I did run the regression model for the largest political parties separately, I did not conduct a thorough analysis of the effect of political affiliation. This is due to the difficulty of quantifying parties along the political spectrum in a way that is most meaningful to this research. This is potentially problematic because there are generally dependencies between party choice and demographic factors. For instance, in urban areas there tend to be more votes for left-wing parties (Voorn, 2021). Fourthly, the variable included to denote list position was categorical, divided into top, middle, and bottom based on list size only. It would have been more accurate to distinguish candidates who were expected to receive a seat based on poll data, but this information was not easily available. Lastly, there is a relatively large proportion of missing data due to omitted postal codes in some of the municipalities. There is no clear geographical clustering of these municipalities, but I cannot confirm the data is missing at random, as the cause is unknown.

The statistical concerns are mainly related to model and test choice. Firstly, because the nonzero values have been separated into a two-part model, the set of results is conditional. Similarly, as the outputs of the two models differ it is impossible to compare the coefficient sizes. The separation appears to be justified as the effect of list position and wealth on same-town residency is contingent on the inclusion of zero values, indicating that the mechanism that generates what candidates receive no votes is slightly different. There are possible methods to avoid this, such as the marginalized two-part Beta regression proposed by Chai et al. (2018) that allows for an examination of the covariates on the overall mean. Secondly, to test for significance I opted to look at the Wald t-values (t-as-z method), with an additional LRT test to verify more unclear cases. These methods are less accurate than more complex techniques for testing significance in mixed models, such as bootstrapping, especially when the sample size is small (Luke, 2017). These methods are more computationally expensive but could yield improvements in precision, although the differences are unlikely to be large considering the used sample size.

In short, while this thesis project suggests the existence of a hometown advantage for candidates in the Dutch Provincial and Water Board elections, there is still room for methodological improvement. Apart from the limitations highlighted above, there are also several other possible avenues that future researchers could take. Prominently, in this research, I attempted to assess the role of political awareness on locality through comparing two elections with both relatively low turnout and interest. Since the differences are only slight, and the two elections occur concurrently, it is perhaps not surprising that differences in regression outcomes are also only slight. Therefore, running a similar analysis for other Dutch elections, such as on the national level could serve as a valuable addition. It would be wise to look at several iterations of the same elections to assess overall differences. Furthermore, just as including political orientation of candidates would have been beneficial, investigating the different role of same-town residence for local and national parties could be a valuable addition. There might be different dynamics among voters who have already chosen for a local political party. These inquiries could help us develop a deeper understanding of the role of candidates' background in the democratic process.

8. BIBLIOGRAPHY

Adam, S., & Maier, M. (2010). Personalization of politics a critical review and agenda for research. *Annals of the International Communication Association*, *34*(1), 213–257.

Ansolabehere, S., & Puy, M. S. (2016). Identity voting. *Public Choice*, *169*(1), 77–95. https://doi.org/10.1007/s11127-016-0371-2

Arzheimer, K., & Evans, J. (2012). Geolocation and voting: Candidate–voter distance effects on party choice in the 2010 UK general election in England. *Political Geography*, *31*(5), 301–310. https://doi.org/10.1016/j.polgeo.2012.04.006

Binnema, H., & Vollaard, H. (2021). The 2019 provincial elections in the Netherlands: The Rise of Forum voor Democratie after a heavily nationalized campaign. *Regional & Federal Studies*, *31*(3), 433–446. https://doi.org/10.1080/13597566.2020.1840365

Boulton, A. J., & Williford, A. (2018). Analyzing Skewed Continuous Outcomes With Many Zeros: A Tutorial for Social Work and Youth Prevention Science Researchers. *Journal of the Society for Social Work and Research*, *9*(4), 721–740. https://doi.org/10.1086/701235

Chai, H., Jiang, H., Lin, L., & Liu, L. (2018). A marginalized two-part Beta regression model for microbiome compositional data. *PLOS Computational Biology*, *14*(7), e1006329. https://doi.org/10.1371/journal.pcbi.1006329

Conover, P. J. (1984). The influence of group identifications on political perception and evaluation. *The Journal of Politics*, *46*(3), 760–785.

de Leeuw, S. (2017). Het geslacht van de kandidaat als heuristisch stemmotief: Een onderzoek naar het effect van politieke sofisticatie en electorale context op gender-based stemgedrag. *Res Publica*, *59*(2), 141–169. https://doi.org/10.5553/RP/048647002017059002001

Ferrari, S., & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, *31*(7), 799–815. https://doi.org/10.1080/0266476042000214501

Fiva, J. H., & Halse, A. H. (2016). Local favoritism in at-large proportional representation systems. *Journal of Public Economics*, *143*, 15–26. https://doi.org/10.1016/j.jpubeco.2016.08.002

Geissinger, E. A., Khoo, C. L. L., Richmond, I. C., Faulkner, S. J. M., & Schneider, D. C. (2022). A case for beta regression in the natural sciences. *Ecosphere*, *13*(2), e3940. https://doi.org/10.1002/ecs2.3940

Gemeente Weert. *Voorlopige uitslag verkiezingen in Weert | Gemeente Weert*. (n.d.). Retrieved July 1, 2023, from https://www.weert.nl/voorlopige-uitslag-verkiezingen-in-weert

Gimpel, J. G., Karnes, K. A., McTague, J., & Pearson-Merkowitz, S. (2008). Distance-decay in the political geography of friends-and-neighbors voting. *Political Geography*, *27*(2), 231–252. https://doi.org/10.1016/j.polgeo.2007.10.005

Górecki, M. A., Bartnicki, S., & Alimowski, M. (2022). Local voting at local elections revisited: ‘Friends and neighbors voting’ at mayoral elections in rural Poland. *Political Geography*, *94*, 102559. https://doi.org/10.1016/j.polgeo.2021.102559

Górecki, M. A., & Marsh, M. (2012). Not just ‘friends and neighbours’: Canvassing, geographic proximity and voter choice. *European Journal of Political Research*, *51*(5), 563–582. https://doi.org/10.1111/j.1475-6765.2011.02008.x

Halm, I. van. (2022, August 16). The Dutch nitrogen crisis shows what happens when policymakers fail to step up. *Energy Monitor*. https://www.energymonitor.ai/policy/the-dutch-nitrogen-crisis-shows-what-happens-when-policymakers-fail-to-step-up/

Havekes, H. J. M., & Elfferich-Rodenburg, I. D. M. (2020). Waterschapsverkiezingen Nieuwe Stijl: Een Succesverhaal? *Tijdschrift Voor Constitutioneel Recht*, 20.

Herron, E. S., & Lynch, M. S. (2019). ‘Friends and Neighbors’ Voting Under Mixed-Member Majoritarian Electoral Systems: Evidence from Lithuania. *Representation*, *55*(1), 81–99. https://doi.org/10.1080/00344893.2019.1581078

Hessing, R. C. (1985). Bij voorkeur: Een onderzoek naar het gebruik van voorkeursstemmen. *Acta Politica*, *20*(2), 157–176.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kavanagh, A., Durkan, W., & D’Arcy, C. (2021). Geographical Factors in Constituency Voting Patterns. In M. Gallagher, M. Marsh, & T. Reidy (Eds.), *How Ireland Voted 2020* (pp. 197–217). Springer International Publishing. https://doi.org/10.1007/978-3-030-66405-3_9

Key, V. O., & Heard, A. (1949). *Southern politics in state and nation* (Vol. 510). New York: AA Knopf.

Kiesraad. (2016a, April 20). *Waterschappen—Verkiezingen—Kiesraad.nl* [Webpagina]. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. https://www.kiesraad.nl/verkiezingen/waterschappen

Kiesraad. (2016b, April 26). *Kieskringen—Verkiezingen—Kiesraad.nl* [Webpagina]. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. https://www.kiesraad.nl/verkiezingen/tweede-kamer/kandidaatstelling/kieskringen

Kiesraad (Rijk). (2023a). *Verkiezingsuitslagen Provinciale Staten 2023* [Data set]. https://data.overheid.nl/dataset/verkiezingsuitslag-provinciale-staten-2023

Kiesraad (Rijk). (2023b). *Verkiezingsuitslagen Waterschappen 2023* [Data set]. https://data.overheid.nl/dataset/verkiezingsuitslagen-waterschappen-2023

Klaassen, N. (2023, March 11). *Peiling: Opmars BBB zet door, CDA halveert, links blok groot in Senaat*. Het Parool. https://www.parool.nl/verkiezingen/peiling-opmars-bbb-zet-door-cda-halveert-links-blok-groot-in-senaat~bc5367db/

Koninkrijksrelaties, M. van B. Z. en. (2011, November 20). *Provinces, municipalities and water authorities—Public administration—Government.nl* [Onderwerp]. Ministerie van Algemene Zaken. https://www.government.nl/topics/public-administration/provinces-municipalities-and-water-authorities

Koninkrijksrelaties, M. van B. Z. en. (2016, February 16). *Provincial tasks—Provinces—Government.nl* [Onderwerp]. Ministerie van Algemene Zaken. https://www.government.nl/topics/provinces/provincial-tasks

Lefevere, J., & Van Aelst, P. (2014). First-order, second-order or third-rate? A comparison of turnout in European, local and national elections in the Netherlands. *Electoral Studies*, *35*, 159–170. https://doi.org/10.1016/j.electstud.2014.06.005

Liu, L., Shih, Y.-C. T., Strawderman, R. L., Zhang, D., Johnson, B. A., & Chai, H. (2019). Statistical Analysis of Zero-Inflated Nonnegative Continuous Data. *Statistical Science*, *34*(2), 253–279.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Marien, S., Dassonneville, R., & Hooghe, M. (2015). How second order are local elections? Voting motives and party preferences in Belgian municipal elections. *Local Government Studies*, *41*(6), 898–916.

Nagtzaam, M. A. (2019). *Second-order electoral personalization. Intra-party preference voting in Belgium and the Netherlands* [PhD Thesis]. Leiden.

Newsom, J. T. (2019). *Significance Testing in Multilevel Regression*.

Partijgeschiedenis. (2021, August 3). Rijksuniversiteit Groningen. https://www.rug.nl/research/dnpp/politieke-partijen/bbb/geschiedenis/

Plutzer, E., & Zipp, J. F. (1996). IDENTITY POLITICS, PARTISANSHIP, AND VOTING FOR WOMEN CANDIDATES. *Public Opinion Quarterly*, *60*(1), 30–57. https://doi.org/10.1086/297738

Provinciale Staten. (n.d.). Parlement.com. Retrieved May 22, 2023, from https://www.parlement.com/id/vibzfyogo7qo/provinciale_staten

Put, G.-J. (2021). Is there a friends-and-neighbors effect for party leaders? *Electoral Studies*, *71*, 102338. https://doi.org/10.1016/j.electstud.2021.102338

Put, G.-J., Smulders, J., & Maddens, B. (2019). How local personal vote-earning attributes affect the aggregate party vote share: Evidence from the Belgian flexible-list PR system (2003–2014). *Politics*, *39*(4), 464–479. https://doi.org/10.1177/0263395718811969

Put, G.-J., von Schoultz, Å., & Isotalo, V. (2020). Fighting over friends and neighbors: The effect of inter-candidate geographic distance on intra-party competition. *Political Geography*, *81*, 102219. https://doi.org/10.1016/j.polgeo.2020.102219

Schelde, A. V. der, & Kanne, P. (2023, March 16). *PS-trends in kaart gebracht*. I&O Research. https://www.ioresearch.nl/actueel/ps-trends-goed-in-kaart-gebracht/

Shugart, M. S., Valdini, M. E., & Suominen, K. (2005). Looking for Locals: Voter Information Demands and Personal Vote-Earning Attributes of Legislators under Proportional Representation. *American Journal of Political Science*, *49*(2), 437–449. https://doi.org/10.1111/j.0092-5853.2005.00133.x

University of Wisconsin-Madison. (n.d.). *Mixed Models: Testing Significance of Effects*. Social Science Computing Cooperative. Retrieved June 12, 2023, from https://www.ssc.wisc.edu/sscc/pubs/MM/MM_TestEffects.html

Van Erkel, P. F. A. (2019). Sharing is caring: The role of voter-candidate similarities in intra-party electoral competition. *European Political Science Review*, *11*(1), 75–89. https://doi.org/10.1017/S175577391800022X

Van Holsteyn, J. J., & Andeweg, R. B. (2012). Tweede orde personalisering: Voorkeurstemmen in Nederland. *Res Publica*, *54*(2).

Vollaard, H., Binnema, H., Denters, B., & Jansen, G. (2021). *Meer zicht op het democratisch middenbestuur: Het kiezersonderzoek in provincies en waterschappen van 2020*. University of Twente. https://doi.org/10.3990/1.9789036552776

von Klipstein, L., Servaas, M. N., Lamers, F., Schoevers, R. A., Wardenaar, K. J., & Riese, H. (2023). Increased affective reactivity among depressed individuals can be explained by floor effects: An experience sampling study. *Journal of Affective Disorders*, *334*, 370–381. https://doi.org/10.1016/j.jad.2023.04.118

Voorn, B. (2021). *De politieke oriëntatie van Nederlandse gemeenten, 2021*.

APPENDICES

APPENDIX 1: DATA SOURCES

Table 6. Data sources descriptions

ID	Dataset	Description	Source	Completeness
1	Candidates provinciale staten/waterschappen (2 files)	For each separate kieskring/waterschap, each candidate running for each political party, with information such as name, gender and place of residence. Originally published in EML format.	Kiesraad (Rijk). (2023). Verkiezingsuitslagen Provinciale Staten 2023 [Data set]. https://data.overheid.nl/dataset/verkiezingsuitslag-provinciale-staten-2023 Kiesraad (Rijk). (2023). Verkiezingsuitslagen Waterschappen 2023 [Data set]. https://data.overheid.nl/dataset/verkiezingsuitslagen-waterschappen-2023	Complete concerning relevant variables
2	Results for each voting station (2 files)	For each voting station the number of votes for each candidate. The candidates are identified by only their list position. Each voting station has an ID, and most have a postal code. Originally published in EML format.		For around 760 stations, the postal code is missing. This is the main cause of missing data in the final datasets.
3	Voting stations	For each voting station, the name, address, postal code, municipality and town, among other variables.	Open State Foundation. (2023). Stembureaus Provinciale Statenverkiezingen 2023 [Data set]. https://ckan.dataplatform.nl/dataset/stembureaus-provinciale-statenverkiezingen-2023	Postal code and address info is missing for around 348 rows. The stations in kieskringen Venlo and Maastricht are incorrectly grouped as one.
4	Shapefile voting stations	A shapefile containing the locations of the voting stations, with name, address, postal code, etc, but no voting station ID. The collection is also slightly different than dataset 3.	ESRI Nederland. (2023). Stembureaus Provinciale Staten en Waterschapsverkiezingen 2023 [Data set]. https://hub.arcgis.com/datasets/esrinl-content::stembureaus-provinciale-staten-en-waterschapsverkiezingen-2023/about	Postal code is missing for around 191 stations, many of which are incorrectly recorded rows.
5	Shapefiles postal codes and municipalities	Shapefiles of the 4-digit postal codes for the Netherlands and of the different municipalities	CBS. (2021). Postcodes- Netherlands [Data set]. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2020	Complete
6	Geographic information per postal code	For each 4-digit postal code information such as resident count, density, average property worth (WOZ), and degree of urbanization (5 categories).	CBS. (2022). Kerncijfers per postcode [Data set]. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode	Nearly complete concerning relevant variables (max 249 out of 4053 missing rows for 1 variable)

APPENDIX 2: DATASET CREATION STEPS
Voting station file

Adding postal code and town-name wherever possible (reduces missing values from 348 to 207)

Loaded datasets 4 and 5 into QGIS and performed a 'join by attributes' transformation to add the 4-digit postal code to all voting stations. Saved the resulting layer as a CSV called 'stembureaus2'

I loaded this file, along with dataset 3 called 'stembureau in R' and performed a left_join operation by the name and municipality of the voting station [3] to add 4-digit postal code and the name of the town as new columns.

If the postal code or town were missing in the voting station file, they were replaced by the 4-digit postal code and town from stembureau2.

Adding kieskring and plaats with a loop

For some rows the kieskring is missing, which we can deduce from the municipality. I looped through the empty kieskring rows and added the kieskring from rows with the same municipality (reduces missing values from 210 to 200).

I did the same for the missing towns, but based on the postal code rather than the municipality (reduces missing values from 346 to 238).

The kieskring in dataset 3 was denoted with a number code. I changed this to words.

Provinciale staten

Joining voting station [3] and results dataset [2] to get location for each.

I created a dataset with every unique postalcode in the voting station file [3]. This is possible because every postal code is only associated with one town, and one kieskring (this is not the case in the waterschappen datasets).

I performed a left_join operation by the 4-digit postal code to attach location information for each voting station in the results dataset [2]. Now, I can see for each station the number of votes that a certain candidate received, and the location of that station.

Joining results [2, 3] and candidate information [1].

I changed the kieskringen names in the candidate dataset [1] to match the results dataset [2,3].

I joined the two datasets by the kieskring, political party, and list position. This resulted in a dataset where we can also see the name and demographic information such as their place-of-residence of each candidate, and how many votes they got in each voting station [1, 2, 3].

Adding geographical information.

Based on the 4-digit postal code, I added geographical information such as average property worth and population size from dataset 6.

Creating variables based on list composition

I created a binary variable for whether the town where the voting station is and the place-of-residence of the candidate match.

I created another binary variable for whether the candidate from that town is the only candidate from that town.

I created a categorical variable for whether a candidate is at the top (top 10%), middle or bottom (bottom 10%) of the party list

I created a binary variable for whether or not a candidate is the last candidate on the list (lijstduwer)

I identified and removed the candidates who are first on the list

Converting raw vote counts to percentages

Pasting together first and last name into one variable

Waterschappen

Adding the waterschap to each voting station in dataset 2.

Because there are instances where one voting station is in multiple waterschappen, I used a different approach for this dataset.

I created a dataset with each waterschap and a political party that is only running in that waterschap. Three waterschappen have no unique party. I left those for later.

Because there are multiple waterschappen in one station, the bureau ID's are not truly unique. I created a new ID based on the structure of the vote results [2].

I joined the waterschap to the results [2], so that each political party that was identified as unique has a value in the waterschap column.

I filled in the waterschap column for all rows in dataset 2 based on the new bureau ID.

I separated the rows where the waterschap column was empty. These belonged to the three waterschappen that had no unique parties, and I identified the parties that are unique among only the three.

I repeated the steps above for this subgroup.

Adding locations

I joined the results dataset [2] with the added waterschappen with the unique voting stations [3] based on the 4-digit postal code.

There are 240059 rows where the town is missing, due to missing postal codes in the result file [2]. This is a similar proportion to the provinciale staten dataset.

Joining the results [2, 3] and the candidates [1] datasets

I changed the names of the waterschappen in the candidate dataset [1] so that they match.

I performed a left_join operation to add the candidate information based on the waterschap, political party, and list position. This resulted in a dataset where we can also see the name and demographic information such as their place-of-residence of each candidate and how many votes they got in each station [1, 2, 3].

Adding geographical information.

Based on the 4-digit postal code, I added geographical information such as average property worth and population size from dataset 6.

Creating variables based on list composition

I created a binary variable for whether the town where the voting station is and the place-of-residence of the candidate match.

I created another binary variable for whether the candidate from that town is the only candidate from that town.

I created a categorical variable for whether a candidate is at the top (top 10%), middle or bottom (bottom 10%) of the party list

I created a binary variable for whether or not a candidate is the last candidate on the list (lijstduwer)

I identified and removed the candidates who are first on the list

Converting raw vote counts to percentages

Pasting together first and last name into one variable

APPENDIX 3: MODEL DIAGNOSTIC CHECK

Table 8 shows that the data structures of each model meet the necessary assumptions for the analyses. The majority of the assumptions are 'met', while some are 'roughly met', meaning that the distributions divert somewhat from the ideal. There is little consensus about what diagnostics are important for beta regression. I follow Geissinger et al. (2022) and investigate the normality of residuals and homoscedasticity. These two are not applicable for the logistic regression and have been marked 'X'. For all models I look at multicollinearity, and the distribution of the random effects (Jaeger, 2008). Logistic regression does not require linearity between independent and dependent variables, but it does require linearity between the predictors and the log odds. There were noticeable outliers for the wealth variable, but after testing the model with and without them, I saw no substantial differences, and decided to leave the data as is. With regards to the beta regression models, there was some visible heteroscedasticity, but none that appeared especially problematic.

Table 7. Assumptions check for each model.

	logPS	**betPS**	**logWB**	**betWB**
Multicollinearity	Met	Met	Met	Met
Linearity	Roughly met	Met	Met	Met
Random effects distribution	Met	Met	Met	Met
Normality of residuals	X	Roughly met	X	Roughly met
Homoscedasticity of residuals	X	Roughly met	X	Roughly met
Outliers	Met	Met	Met	Met

APPENDIX 4: CHECKS FOR DIFFERENT POLITICAL PARTIES AND UNITS

Table 8. Estimate size of same town residence for 10 biggest parties in the Provincial and Water Board elections

Party	Coefficient PS	Coefficient WS
BBB	0.529***	0.365***
CDA	0.429***	0.447***
D66	0.250***	
GroenLinks	0.360***	
JA21	0.297***	
PvDA	0.442***	0.390***
PvDD	0.330***	0.236***
PVV	0.228***	
SP	0.410***	
VVD	0.403***	0.336***
WaterNatuurlijk		0.372***
AWP		0.256***
ChristenUnie	0.320***	0.425***
50PLUS	0.259***	0.344***
SGP	0.489***	0.652***

Table 9. Estimate size of same town residence for each kieskring and Water Board

Kieskring	Coefficient size	Waterschap	Coefficient size
Drenthe	0.465***	**Waterschap Scheldestromen**	0.557***
Flevoland	0.379***	**Waterschap Rivierenland**	0.539***
Fryslan	0.403***	**Waterschap Rijn en IJssel**	0.470***
Arnhem	0.578***	**Waterschap Limburg**	0.557***
Nijmegen	0.574***	**Waterschap Hollandse Delta**	0.400***
Groningen	0.235***	**Waterschap Drents Overijsselse Delta**	0.554***
Amsterdam	0.208***	**Waterschap De Dommel**	0.710***
Den Helder	0.621***	**Vallei & Veluwe**	0.553***
Haarlem	0.509***	**Schieland en de Krimpenerwaard**	0.487***

Overijssel	0.519***	**HHS van Delfland**	0.260***
Maastricht	0.411***	**HH van Rijnland**	0.565***
s-Hertogenbosch	0.660***	**HH Hollands Noorderkwartier**	0.561***
Tilburg	0.584***	**De Stichtse Rijnlanden**	0.318***
Utrecht	0.307***	**Brabantse Delta**	0.464***
Zeeland	0.518***	**Amstel, Gooi en Vecht**	0.230***
Dordrecht	0.630***	**Aa en Maas**	0.570***
Leiden	0.629***	**Waterschap Zuiderzeeland**	0.388***
Rotterdam	0.355***	**Vechtstromen**	0.487***
		Wetterskip Fryslân	0.398***
		Waterschap Hunze en Aa's	0.362***
		Waterschap Noorderzijlvest	0.114***