Implied Comparative Advantage in Electoral Geography:
Recovering the Latent Structure of Candidate-Location Matching

Gabriel C. Caseiro*

July 18, 2026

Abstract

This paper introduces the implied comparative advantage framework to the study of electoral geography. Adapting the methodology developed by Hausmann et al. (2022) for economic geography, I recover the latent structure of candidate-location matching from the full cross-sectional pattern of vote intensities in Brazilian federal deputy elections. The approach constructs similarity matrices across municipalities and across candidates, and uses them to estimate an implied vote intensity for every candidate-municipality pair — a counterfactual prediction of how well each candidate should perform in each municipality given the general structure of political competition in the state. Using data from all 26 Brazilian states across seven elections (1998–2022), I find that: (1) the implied measures explain approximately half the variance in observed vote intensity; (2) the residuals from this cross-sectional fit predict future changes in electoral geography, with candidates converging toward their implied vote geography over time; (3) this convergence strengthens monotonically with the time horizon, with 1998 residuals predicting vote geography changes through 2022 more strongly than over a single electoral cycle; and (4) the relative weight of municipality-based and candidate-based information shifts over time, with candidate similarity becoming increasingly informative — consistent with a flattening of the spatial structure of political information. The framework produces objects new to electoral geography — dyadic predictions, similarity structures, and residuals that decompose vote geography into structural and idiosyncratic components — and demonstrates that a stable latent structure of political demand and supply underlies Brazilian electoral geography.

*New York University, New York, NY, USA; Email: gcc9360@nyu.edu

1 Introduction

What determines the spatial distribution of a candidate's votes? Electoral geography has developed a rich tradition of characterizing vote maps — measuring how concentrated or dispersed a candidate's support is, whether high-vote areas cluster geographically, and how these patterns vary across electoral systems and institutional settings (Forest, 2018; Ames, 1995). This paper asks a complementary question: what does the full cross-sectional pattern of vote intensities — across all candidates and all municipalities simultaneously — reveal about the latent structure of political competition in an electoral district? And can this structure tell us something about where a candidate's electoral geography is headed?

To address these questions, I draw on the concept of implied comparative advantage developed by Hausmann et al. (2022) to study the relationship between industries and locations in economic geography. I adapt their methodology to recover the latent structure of candidate-municipality matching from the observable pattern of votes. The approach treats each state-election as a system in which multiple candidates compete across multiple municipalities, and uses the full matrix of vote intensities to estimate how similar any two municipalities are (based on which candidates perform well in each) and how similar any two candidates are (based on where each performs well). From these similarity structures, it constructs a counterfactual prediction — the implied vote intensity — for every candidate-municipality pair, including pairs where the candidate currently receives no votes.

I apply this framework to all Brazilian states across seven elections for the federal Chamber of Deputies (1998–2022), producing three main findings. First, the implied vote intensity measures explain a substantial share of observed vote intensity. Across 182 state-elections, the median R^2 from the cross-sectional regression of observed on implied vote intensity is approximately 0.50 for all candidates (and 0.57 for competitive candidates). This demonstrates that a recoverable latent structure underlies the spatial distribution of votes — a structure shaped by the match between candidate attributes and municipality characteristics that is not directly observed but can be inferred from the aggregate pattern.

Second, the residuals from this cross-sectional fit — the gap between observed and implied vote intensity — predict future changes in electoral geography. Candidates who over-perform their structural prediction in one election tend to lose vote intensity in subsequent elections; those who under-perform tend to gain. This convergence holds across all 26 states, in every election pair examined, and — critically — strengthens with the time horizon. Residuals from 1998 predict vote geography changes through 2022 more strongly than they predict changes over a single electoral cycle. This rules out simple mean reversion and suggests that the implied measures capture slow-moving structural forces: a latent electoral potential that gradually asserts itself over decades, despite the turbulence of individual elections, partisan realignments, and changing candidate fields.[1]

[1]Because this long-run convergence may be driven by distinct strategic paths, I extend the core framework in Appendix A to evaluate the underlying political economy mechanisms. Specifically, I leverage variations in supply-side legislative resources (incumbency status) and localized information shocks (the rollout of 3G mobile broadband) to test whether actor-centered incentives or shifting informational environments govern this convergence result.

Third, the relative weight of candidate-based and municipality-based information in predicting vote intensity varies systematically across states and shifts over time. In most states, knowing which municipalities are politically similar has become less informative over the period, while knowing which candidates share similar profiles has become more informative. This temporal shift is consistent with the hypothesis that the spatial structure of political information has flattened — as digital media and nationalized party brands erode the distinctiveness of place-based political environments, the electoral market becomes increasingly segmented by candidate type rather than by municipal characteristics.

The contribution of this paper is primarily methodological, with substantive implications. The implied comparative advantage framework provides electoral geography with a tool that operates at the candidate-municipality level, producing three objects that are new to the literature: a measure of implied vote intensity for every candidate-municipality pair (what the latent political structure predicts), a residual capturing the idiosyncratic deviation from that prediction, and similarity matrices that reveal the latent organization of political competition across municipalities and across candidates. The convergence results demonstrate that this decomposition is not merely descriptive: the idiosyncratic component contains real information about the future trajectory of electoral geography.

This paper engages with several strands of literature. In the Brazilian electoral geography tradition, Ames (1995, 2001) established the foundational connection between the spatial distribution of a deputy's votes and their subsequent legislative behavior — showing that electoral geography matters for representation. Carvalho (2003) extended this by demonstrating that deputies behave in congress according to their spatial voting type. Avelino et al. (2011, 2016) and Silva and Silotto (2018) brought measurement rigor to the study of vote concentration, developing indices that track its evolution across states and elections. These contributions share a focus on characterizing *individual candidates' vote distributions* and connecting those distributions to political outcomes. A more recent strand, represented by Guarnieri and da Silva (2022), has shifted attention to the *generative processes* behind vote geography — modeling how information diffuses through spatial networks and how parties strategically regionalize their candidate lists to manage intra-party competition. This paper is closer in spirit to this second strand: it seeks to uncover the structure beneath observed vote maps, though through a different method — recovering latent patterns from the cross-sectional correlation structure of the full candidate $\times$ municipality matrix rather than modeling a specific mechanism like information diffusion or party strategy.

More broadly, the paper connects to a longstanding tension in electoral geography between compositional and contextual explanations for spatial variation in voting (Agnew, 1996; King, 1996; Azevedo, 2023). Compositional explanations attribute spatial differences to the characteristics of people who live in different places; contextual explanations attribute them to features of the places themselves. The implied comparative advantage framework does not require the researcher to choose between these accounts. The municipality similarity matrix recovers political similarity from the aggregate pattern of voting without specifying whether the underlying drivers are the socioeconomic composition of the electorate, the information environment, the organizational infrastructure of parties, or some other place-based feature.

Finally, the methodological approach draws on the economic complexity literature. Hausmann et al. (2022) showed that hidden information about inter-industry and inter-location relatedness can be recovered from the cross-sectional pattern of production, and that the resulting measure of "implied comparative advantage" predicts long-run structural change. The product space framework of Hidalgo et al. (2007) and Hidalgo and Hausmann (2009) demonstrated that the same relatedness structure shapes the evolution of countries' export baskets. I adapt these tools to the electoral setting, where candidates play the role of industries and municipalities play the role of locations. The theoretical translation rests on a premise well established in political science — that vote intensity reflects a match between candidate attributes and voter preferences mediated by information environments — which I develop formally in Section 2.

The paper proceeds as follows. Section 2 develops the theoretical framework. Section 3 describes the data and empirical implementation. Section 4 presents the main results. Section 5 discusses implications and limitations. Section 6 concludes.

2 Theoretical Framework

2.1 The Structure of Political Competition in Open-List Systems

Brazil's open-list proportional representation system creates a distinctive electoral geography problem. Since voters choose individual candidates rather than parties, and electoral districts correspond to entire states (some containing hundreds of municipalities and electing up to 70 deputies) each election produces a dense matrix of vote intensities linking every candidate to every municipality. This matrix is the raw material for the study of electoral geography in Brazil, and its spatial structure has been the object of sustained scholarly attention.

The tradition developed in works such as Ames (2001) and Avelino et al. (2016) treats this matrix column by column — characterizing each candidate's vote distribution and linking its spatial properties to political behavior. Mayhew (1974)'s electoral connection, originally theorized for single-member districts, finds its analogue in the emergence of "informal districts" under OLPR: because campaigns are costly and districts are vast, candidates concentrate their efforts in specific territories (Ames, 2001; Silva and Silotto, 2018). The resulting vote geography shapes parliamentary action: concentrated-dominant deputies direct pork-barrel resources to their strongholds, while dispersed deputies pursue more thematic-corporate strategies (Carvalho, 2003; Rocha, 2021).

Yet this column-by-column approach — analyzing each candidate's vote map individually — leaves unexploited the information contained in the *relationships between columns*. When two candidates exhibit similar spatial patterns of support across municipalities, this reveals something about both the candidates (they appeal to similar electorates) and the municipalities (they share political characteristics that make them responsive to the same kind of appeal). When two municipalities show correlated vote intensity profiles across many candidates, this reveals a latent political similarity that may or may not reduce to geographic proximity, socioeconomic composition, or shared media markets. The implied comparative advantage framework exploits precisely this relational structure. Rather than summarizing each candidate's vote map in isolation, it uses the full matrix to recover the hidden patterns of similarity among candidates and among municipalities.

But why should such relational structure exist in the first place? The answer lies in the incentives of the two sets of actors who produce the vote matrix: voters and candidates. The structure is, thus, a consequence of how heterogeneous voters choose among heterogeneous candidates in an environment where political information is spatially organized.

The voter's problem

Under OLPR, a voter must select a single candidate from a potentially long list of names — in large Brazilian states, hundreds of candidates across party lists. This information-intensive choice involves two conceptually distinct dimensions (Shugart et al., 2005; Jankowski, 2016). The first is a *preference* dimension: what attributes does the voter value in a representative? Voters, for instance, differ in the weight they place on local representational capacity (i.e., the expectation that a deputy will channel resources to their municipality and serve as a territorial interlocutor with the federal government) versus broad candidate quality: ideology, policy positions, competence, or national influence (Downs, 1957; Stokes, 1963). The second is an *information* dimension: how precisely can the voter evaluate each candidate along the attributes they care about? A voter can only act on their preferences to the extent that they can assess candidates, and this assessment capacity is not spatially uniform — it depends on the technology of information transmission connecting the candidate to the voter (Lupia and McCubbins, 1998).

For candidates whose careers are rooted in local office, family networks, or regional media, the information signal is spatially structured: strong where the candidate is personally known and decaying with geographic distance (Key, 1949; Gimpel et al., 2008; Guarnieri and da Silva, 2022). For candidates who enter politics through national media, digital platforms, or ideological movements, the information signal is spatially flatter: accessible to voters across the state with roughly comparable precision. The interaction between preferences and information contributes the spatial patterns we observe. A locally rooted candidate is expected to produce high vote intensity in their home region through two reinforcing channels: voters who value local representation find a direct preference match, *and* voters of all types have richer information about this candidate than about most alternatives — a proximity-based informational advantage that translates into votes even from voters who might otherwise prioritize broad quality (Shugart et al., 2005; Bräuninger et al., 2024). Far from the candidate's base, both channels should weaken simultaneously, producing steep spatial decay. A nationally known candidate, by contrast, is expected to produce a flatter vote intensity surface: only the geographic distribution of preference types generates spatial variation, since the informational channel that powerfully reinforces concentration for local candidates is absent.

The candidate's problem

The match structure is not only demand-driven: it also reflects supply-side strategic choices. Candidates decide how to position themselves in the political landscape: which career path to pursue, which party to join, which communication technology to invest in, and where to allocate campaign resources (Ames, 2001; Carey and Shugart, 1995). These choices are shaped by the candidate's comparative advantage — their pre-existing networks, name recognition, and media access — and, crucially, by the incentive structure of the electoral system itself.

Multi-seat competition in large-magnitude districts creates centrifugal incentives. Cox (1990) shows that as the number of seats per district increases, the effective vote threshold for winning a seat decreases, and candidates are pushed toward differentiated rather than convergent positions — they can win by appealing to a segment of the electorate rather than by assembling a broad majority. Myerson (1993) formalizes a related logic: in multi-seat elections where candidates can target resources, the equilibrium involves cultivating "favored minorities" — concentrating promises on narrow groups rather than spreading them broadly. In the Brazilian context, where district magnitudes range from 8 to 70, these centrifugal forces are strong. They manifest as the geographic differentiation that Ames (2001) documented: candidates build *redutos eleitorais* (i.e., electorally defined territories where they concentrate campaign effort, constituency service, and distributive resources) because the electoral system rewards niche-building over broad-based appeal.

Intra-party competition under OLPR adds a second layer to this centrifugal pressure (Carey and Shugart, 1995; Cheibub and Nalepa, 2020). Because candidates on the same list compete with each other for personal votes, co-partisans face a coordination problem: the party's collective interest is to cover the state's political geography efficiently, filling territorial niches without costly overlap, but each individual candidate prefers to target the most electorally productive municipalities even when this means competing with a running mate. Silotto (2019) and Put et al. (2020) document how parties and candidates respond by regionalizing their efforts, effectively partitioning the state's political geography among co-partisans. When coordination succeeds, the party list produces a set of candidates with complementary vote geographies — each occupying a distinct position in the latent space. When it fails, overlapping candidates split votes in municipalities where their profiles are similarly well-matched, and each receives fewer votes than their structural match quality would predict.

This coordination logic explains why the candidate similarity matrix contains substantive information. Two candidates are similar not because they arbitrarily ended up with correlated vote maps, but because they offer similar bundles of attributes that appeal to similar voter types. The similarity structure reflects the underlying organization of the electoral market into niches defined by the interaction of voter demand and candidate supply.

Why municipalities are politically similar

The municipality similarity matrix recovers the latent political similarity between two places. From the perspective developed above, two municipalities are politically similar when they share features that make their voters responsive to the same kinds of candidates. These features include socioeconomic composition (municipalities with similar economic bases and urbanization levels tend to have similar distributions of voter preferences), information environments (municipalities in the same media market or with similar digital penetration expose their voters to similar candidates), party organizational infrastructure (where the same parties are locally organized, the same candidates benefit from organizational support), and position in the urban hierarchy (regional centers, metropolitan peripheries, and small rural towns tend to face distinct political dynamics that produce distinct patterns of candidate support; see Guarnieri and da Silva 2022).

Geographic proximity correlates with many of these features, which is why the municipality similarity matrix is expected to exhibit spatial decay. But proximity is neither necessary nor sufficient: distant municipalities can share socioeconomic profiles, and adjacent municipalities can differ sharply if they belong to different media markets or have different party structures. The framework recovers this richer similarity structure without requiring the researcher to specify or measure any of these underlying features — addressing, in practice, the tension between compositional and contextual accounts of spatial voting that Azevedo (2023) identifies as a central challenge for electoral geography.

2.2 Formalizing the Latent Space

The previous subsection established that the interaction between heterogeneous voter preferences, spatially structured information, and candidates' strategic positioning generates systematic patterns in the candidate × municipality vote matrix. We now formalize this intuition following the framework of Hausmann et al. (2022), that shows how the latent structure can be recovered from observable data and used to construct counterfactual predictions.

Let the attributes of candidate c be represented by a parameter ψ_c in a latent attribute space $\mathbb{S}$, and the characteristics of municipality l by a parameter λ_l in the same space. As argued above, ψ_c encodes the type of match a candidate offers (i.e., how their career background, communication technology, party brand, and policy positions interact with the preference and information dimensions of voter choice). The parameter λ_l encodes the type of demand a municipality expresses (i.e., the distribution of voter preferences, the information environment, and the organizational infrastructure of local politics). Neither ψ_c nor λ_l is directly observed by the researcher, but the micro-level argument implies that vote intensity is a decreasing function of the distance between them:

$$R_{cl} = f(d(\psi_c, \lambda_l)) \tag{1}$$

where d is a distance metric on $\mathbb{S}$ and f is strictly decreasing: $f(0) = 1$ (perfect match yields maximum intensity) and $f(d_{\max}) = 0$ (maximum mismatch yields zero intensity). This formalization nests the spatial voting tradition, in which voters prefer candidates closer to their ideal point in a latent ideological space (Downs, 1957; Enelow and Hinich, 1989; Poole and Rosenthal, 1985), but generalizes it: the latent space need not be ideological, and the "distance" captures not only preference misalignment but also information barriers — the two channels developed in the previous subsection.

The implied comparative advantage approach differs from ideal point estimation (Clinton et al., 2004) in an important respect. Rather than estimating the latent positions of actors in a pre-specified low-dimensional space, it recovers the *similarity structure* between actors from the aggregate pattern of outcomes, without specifying the dimensionality or content of the latent space. The researcher need not decide whether the relevant dimensions are ideological, geographic, partisan, or socioeconomic — the correlations in the data reveal which candidates are functionally similar and which municipalities are politically similar, whatever the underlying reasons. This agnosticism is a strength in the Brazilian context, where the forces shaping vote geography (such as local networks, media markets, party machines, ideological movements) are heterogeneous across states and changing over time.

2.3 From Latent Structure to Observable Similarity

The key analytical step is showing that the hidden structure can be recovered from observable vote patterns. The logic, following Hausmann et al. (2022), is as follows.

If two municipalities l and l' have similar latent attributes ($\lambda_l \approx \lambda_{l'}$), then by the match function above, they should exhibit similar patterns of vote intensity across candidates: the same candidates should tend to over-perform and under-perform in both places. The correlation between their vote intensity vectors is therefore an observable proxy for the distance between their latent attributes. The municipality similarity index is:

$$\phi_{ll'} = (1 + \text{corr}\{r_l, r_{l'}\})/2 \tag{2}$$

where r_l is the vector of candidate vote intensities in municipality l. This ranges from 0 (opposite political profiles) to 1 (identical profiles), with 0.5 indicating no systematic relationship. The candidate similarity index $\phi_{cc'}$ is defined symmetrically.

$$\phi_{cc'} = (1 + \text{corr}\{r_c, r_{c'}\})/2 \tag{3}$$

Hausmann et al. (2022) prove that under the model assumptions, these similarity measures strictly decrease as the distance between latent parameters increases. Without knowing what ψ and λ are, we can infer the distance between them from observable correlations. This result holds for general distance metrics and functional forms, not just for specific parametric choices.[2]

2.4 Implied Vote Intensity

With the similarity matrices in hand, we construct implied vote intensity measures for each candidate-municipality pair. The municipality-space density estimates the implied intensity of candidate c in municipality l as a weighted average of c's performance in the k most similar municipalities:

[2] See Appendix A.1.1 of Hausmann et al. (2022).

$$\hat{R}_{cl}^{[L]} = \sum_{l' \in L_l(k)} \frac{\phi_{ll'}}{\sum_{l'' \in L_l(k)} \phi_{ll''}} R_{cl'} \tag{4}$$

The candidate-space density estimates it as a weighted average of the most similar candidates' performance in municipality l:

$$\hat{R}_{cl}^{[C]} = \sum_{c' \in C_c(k)} \frac{\phi_{cc'}}{\sum_{c'' \in C_c(k)} \phi_{cc''}} R_{c'l} \tag{5}$$

Neither measure uses candidate c's own performance in municipality l as an input. The implied vote intensity is constructed entirely from information about similar candidates or similar municipalities. This means implied values can be computed even for candidate-municipality pairs with zero observed votes — a feature that is crucial for the extensive margin analysis.

The two density measures capture potentially different information. Municipality-space density answers: "given how this candidate performs in similar municipalities, what should we expect here?" Candidate-space density answers: "given how similar candidates perform in this municipality, what should we expect?" Their relative explanatory power reveals whether the electoral market in a given state is primarily segmented by place (municipality-space dominates) or by candidate type (candidate-space dominates).

2.5 Disturbances, Strategic Adjustment, and Convergence

In any given election, observed vote intensity deviates from the latent match quality. A candidate may not yet have penetrated a well-matched municipality due to information barriers; a coordination failure among co-partisans may suppress a well-matched candidate's support; an unusually effective campaign may inflate a poorly-matched candidate's performance. These deviations — the disturbance term ε_{cl} — mean we observe $\tilde{R}_{cl} = R_{cl} + \varepsilon_{cl}$ rather than the "true" match-determined intensity.

The central dynamic prediction is convergence: over time, these disturbances diminish. This can be understood as a process of strategic adjustment toward equilibrium. The "true" comparative advantage of a candidate-municipality pair represents the vote intensity that would obtain if all actors had full information and coordination were costless. In practice, this equilibrium is never fully realized. Information frictions prevent voters from discovering well-matched candidates. Coordination failures lead parties to misallocate candidates across territories. Campaign resource constraints force candidates to leave structurally promising municipalities untargeted. Convergence reflects the gradual resolution of these frictions as actors learn, adapt, and adjust their strategies over successive elections.[3]

The residuals from the cross-sectional regression capture these frictions: they measure

[3] The microfoundations developed above suggest several mechanisms through which convergence operates (i.e., information diffusion, strategic targeting by incumbents, competitive displacement, and party learning) each rooted in actor incentives and each generating distinct comparative statics about when convergence should be stronger or weaker. These mechanisms and their testable implications are developed and explored empirically in Appendix A.

the deviation of the observed electoral structure from the equilibrium implied by the latent match quality. If we denote the residual from the cross-sectional regression as $\hat{\varepsilon}_{cl,t_0}$, the convergence hypothesis predicts:

$$\tilde{R}_{cl,t_1} - \tilde{R}_{cl,t_0} = \beta \cdot \hat{\varepsilon}_{cl,t_0}, \quad \beta < 0 \tag{6}$$

A negative β means candidates who over-performed their structural prediction tend to lose vote share subsequently, while those who under-performed tend to gain. The theory further predicts that $|\beta|$ should increase with the time horizon: frictions are partially resolved in each electoral cycle, but full adjustment takes multiple cycles because learning is incremental, coordination is imperfect, and institutional rigidities dissolve slowly.

This prediction also extends to the extensive margin. A candidate with high implied vote intensity but zero observed votes in a municipality is "unexpectedly absent" — and should be more likely to appear there in future elections. Conversely, a candidate with low implied intensity but positive observed votes is "unexpectedly present" — and should be more likely to disappear.

2.6 Summary of Hypotheses

The framework presented above generates four testable predictions:

- **H1 (Cross-sectional fit).** The implied vote intensity measures — candidate-space and municipality-space density — should be strongly and positively associated with observed vote intensity.
- **H2 (Convergence).** The residual between observed and implied vote intensity in election t_0 should be negatively associated with the change in observed vote intensity between t_0 and t_1.
- **H3 (Time horizon).** The convergence relationship should strengthen as the time horizon $(t_1 - t_0)$ increases.
- **H4 (Extensive margin).** The implied measures should predict which currently absent candidate-municipality pairs will appear, and which currently present pairs will disappear, in future elections.

3 Data and Implementation

3.1 Data

The analysis uses municipality-level vote counts for all candidates running for the Brazilian federal Chamber of Deputies (*Câmara dos Deputados*) across seven elections (1998, 2002, 2006, 2010, 2014, 2018, and 2022) and all 26 states[4]. The core electoral data come from

[4]Because the Federal District (DF) constains a single municipality (Brasília), it is not considered in this analysis.

Brazil's *Tribunal Superior Eleitoral* (TSE), which provides vote returns at the polling-booth level (*voto-seção*) that I aggregate ate the municipality level. The TSE data also include candidate-level information (*consulta_candidato*), such as party affiliation, coalition membership, electoral number, and candidate identifier.

Brazil's open-list proportional representation system makes this setting particularly well-suited for the implied comparative advantage framework. Voters choose individual candidates from party lists, producing a rich cross-sectional pattern of vote intensities across municipalities. Each state elects between 8 and 70 deputies (depending on population), with candidate fields often exceeding several hundred names. This yields matrices with substantial variation on both dimensions — enough candidates to compute meaningful municipality similarity, and enough municipalities to compute meaningful candidate similarity. Figure 1 reports the number of municipalities and the mean number of candidates (across 1998-2022) in each Brazilian states.

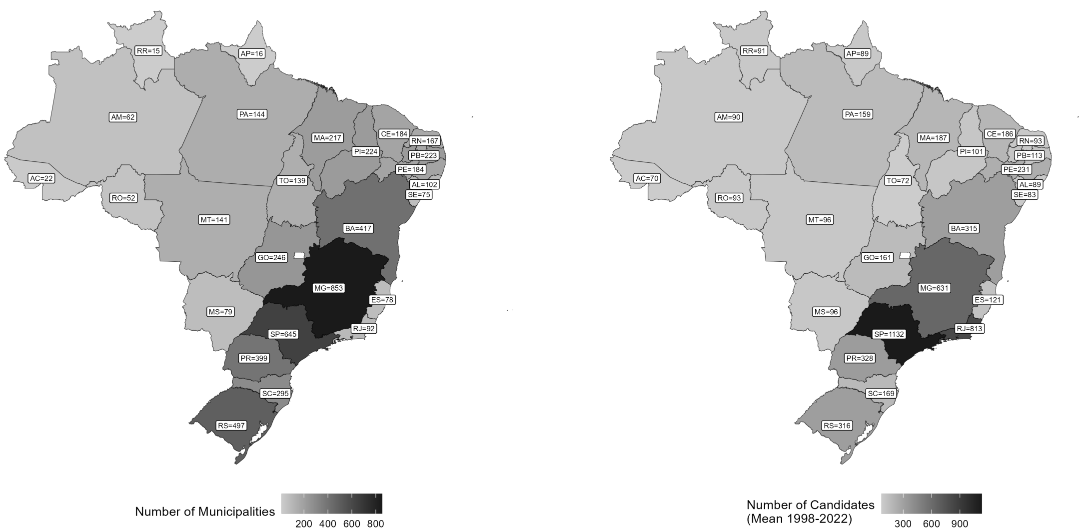

Figure 1: Number of Municipalities and Candidates across States

3.2 Constructing the Location Quotient

For each state-election, I normalize vote counts to construct a location quotient (LQ) that captures the relative intensity of each candidate-municipality pair:

$$R_{cl} = \frac{V_{cm}/V_c}{E_m/E_s} \tag{7}$$

where V_{cm} is the number of votes for candidate c in municipality m, $V_c = \sum_m V_{cm}$ is the candidate's total votes in the state, E_m is the total number of votes in municipality m, and $E_s = \sum_m E_m$ is the state total. An LQ of 1 means the candidate receives exactly the vote intensity expected given the municipality's share of state voters; values above 1 indicate over-representation, and values below 1 indicate under-representation. This measure is analogous to the Revealed per-Capita Advantage (RpCA) used by Hausmann et al. (2022), and is closely related to the standard Location Quotient used in regional economics (e.g., Panzera et al., 2022).

Handling small-municipality instability

Location quotients are known to be unstable in small areas (O'Donoghue and Gleave, 2004). When a municipality's share of the state electorate is very small, even a handful of votes can produce an extreme LQ. A municipality with 50 voters where the candidate receives 5 votes could generate an LQ of 15 or more if the candidate's statewide share is modest. These extreme values are arithmetic artifacts of small denominators rather than signals of genuine geographic concentration, and they can distort the implied compartaive advantage calculation.

I address this problem by capping the LQ at $1/x_{max}$, where x_{max} is the electorate share of the state's largest municipality.[5] The logic is that $1/x_{max}$ is the theoretical maximum LQ that the largest municipality could achieve (if all of a candidate's votes came from that municipality), so capping at this value imposes a natural symmetry on the scale. In practice, this truncates only extreme values in very small municipalities.

3.3 Similarity and Density Construction

For each state-election, I compute the municipality similarity matrix $\phi_{ll'}$ and candidate similarity matrix $\phi_{cc'}$ as scaled Pearson correlations of LQ vectors, following equations 2 and 3. The candidate-space density $\hat{R}_{cl}^{[C]}$ and municipality-space density $\hat{R}_{cl}^{[L]}$ are computed as weighted averages of the k most similar comparators, with the similarity values serving as weights (as defined in equantions 4 and 5). Following the rule of thumb established by Duda (2001) and adopted by Hausmann et al. (2022), I set $k = \lfloor \sqrt{N} \rfloor$, where N is the number of candidates (for candidate-space density) or municipalities (for municipality-space density) in the state-election.

3.4 Regression Specifications

Cross-sectional fit

For each state-election, I estimate:

$$R_{cl} = \alpha + \beta_C \hat{R}_{cl}^{[C]} + \beta_L \hat{R}_{cl}^{[L]} + \varepsilon_{cl} \tag{8}$$

with standard errors clustered at the candidate level. The regression is estimated in levels rather than logs because many candidate-municipality pairs have zero votes, precluding a log transformation without arbitrary imputation. I report results for both the full sample of

[5] Hausmann et al. (2022) uses a similar procedure, ceiling the RpCA by the highest possible value for the most populous country in the world.

candidates and a restricted sample of *competitive candidates*, defined as those receiving at least half the votes of the least-voted elected deputy in the state-election.

Convergence

For each pair of elections (t_0, t_1), I identify candidates who ran in both elections (matched on candidate identifiers from the TSE registry) and estimate:

$$\frac{R_{cl,t_1} - R_{cl,t_0}}{(t_1 - t_0)/4} = \alpha + \beta_\varepsilon \hat{\varepsilon}_{cl,t_0} + e_{cl} \tag{9}$$

where $\hat{\varepsilon}_{cl,t_0}$ is the residual from the cross-sectional regression in t_0. The convergence hypothesis predicts $\beta_\varepsilon < 0$: candidates who over-perform their structural prediction in t_0 tend to lose vote intensity subsequently, while those who under-perform tend to gain.

Extensive margin

For each pair of consecutive elections, I define candidate-municipality presence using LQ thresholds: a pair is *absent* if $R_{cl} = 0$ and *present* if $R_{cl} > 0$. An *appearance* is a transition from absent to present; a *disappearance* is a transition from present to absent. Using a linear probability model, I first regress presence in t_0 on the density measures to obtain a residual, then test whether this residual predicts appearances (among initially absent pairs) and disappearances (among initially present pairs).

4 Results

4.1 The Latent Political Structure

Similarity distributions. Figure 2 shows the distribution of municipality similarity and candidate similarity values for São Paulo, Brazil's most populous state, in 2022. The full distribution of municipality similarities is roughly symmetric with a peak slightly above 0.5 (median ≈ 0.54), indicating modest political differentiation across municipalities — consistent with the "noisy world" prediction of the Hausmann et al. (2022) simulation. Restricting to each municipality's most similar comparator or its top k comparators shifts the distribution rightward, showing that the nearest pairs are substantially more similar than the average pair. Candidate similarities show a qualitatively similar pattern, with a median around 0.51 and a pronounced right tail of highly similar candidate pairs when restricting to the most similar or top k comparators.

What municipality similarity captures. Figure 3 maps the municipality similarity values for four focal municipalities in São Paulo — Araraquara, Bauru, Santos, and São Paulo capital — showing how similar each other municipality is to the focal one. Similarity radiates outward from each focal municipality but does not reduce to geographic contiguity: the pattern follows regional political structure rather than simple distance decay. This is confirmed by Figure 4, which plots municipality similarity against pairwise geographic distance for all municipality pairs in SP 2022. The relationship is clearly negative — nearby municipalities

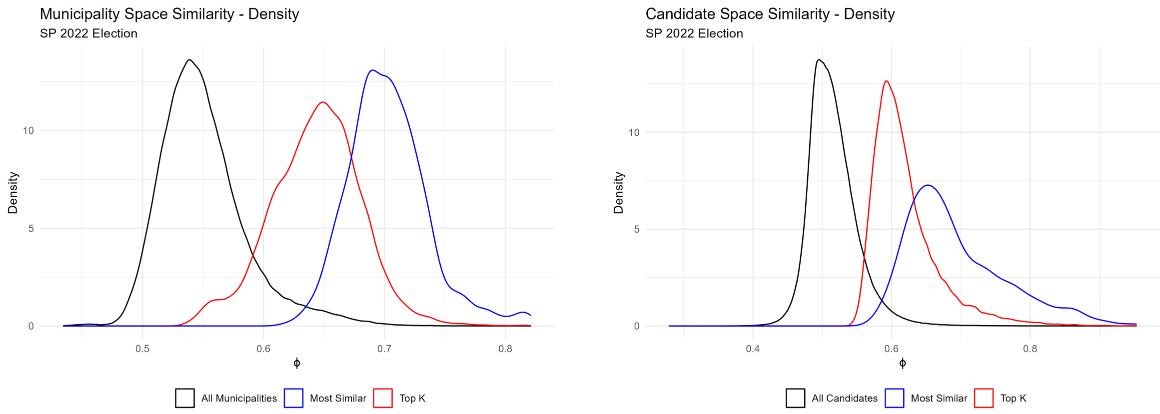

Figure 2: Municipality and Candidate Space Densities

tend to be more politically similar — but with substantial dispersion at every distance, indicating that political similarity captures more than physical proximity. Table B.1 and figure 12 in the Appendix show that all top pairs are geographically proximate and within the same regions.

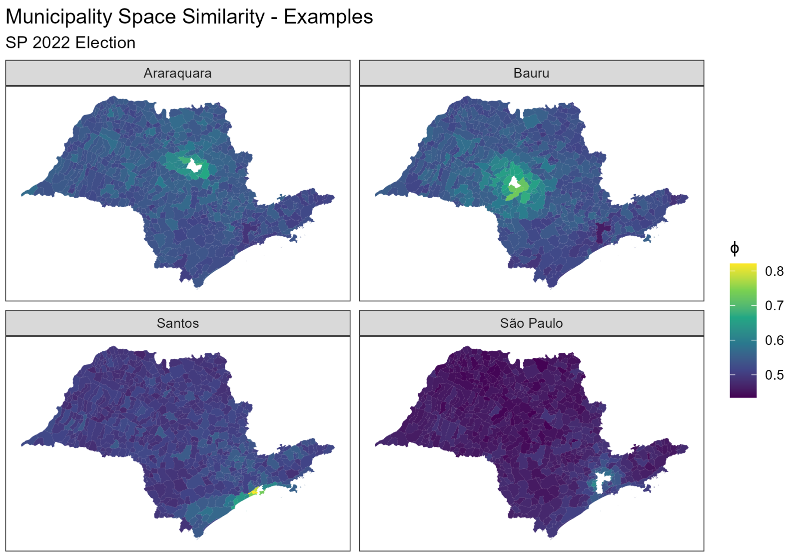

Figure 3: Municipality Similarity — Examples

What candidate similarity captures. Table 1 shows the most similar pairs among elected candidates in SP 2022. Two patterns stand out. First, concentrated candidates with overlapping geographic bases appear as similar pairs: Rosana Valle and Paulo Barbosa ($\phi = 0.890$), both based along the São Paulo coast, have nearly identical spatial footprints (shown in Figure 5). Second, dispersed ideological candidates cluster together: Marina Silva and Erika Hilton ($\phi = 0.843$), Guilherme Boulos and Kim Kataguiri ($\phi = 0.817$) — despite coming from opposite ends of the ideological spectrum — have similar *spatial structures* (both with low vote concentration and diffuse statewide support). The similarity measure captures commonality in the geography of support, not in ideology per se. Indeed, across the most similar pairs, only one has candidates from the same party.

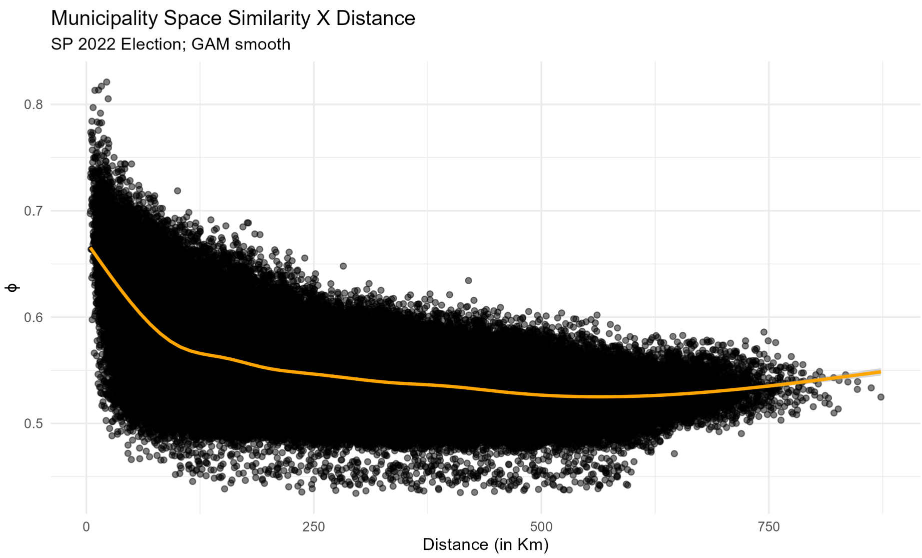

Figure 4: Municipality Similarity X Distance

	Candidate i				Candidate i'			
ϕ	Number	Name	Votes	Gini	Number	Name	Votes	Gini
0.890	2227	ROSANA DE OLIVEIRA VALLE	216437	0.443	4533	PAULO ALEXANDRE PEREIRA BARBOSA	170378	0.376
0.872	1599	ALBERTO PEREIRA MOURÃO	114234	0.451	2227	ROSANA DE OLIVEIRA VALLE	216437	0.443
0.843	1818	MARIA OSMARINA MARINA DA SILVA VAZ DE LIMA	237526	0.171	5070	ERIKA SANTOS SILVA	256903	0.172
0.843	1818	MARIA OSMARINA MARINA DA SILVA VAZ DE LIMA	237526	0.171	5010	GUILHERME CASTRO BOULOS	1001472	0.145
0.841	5010	GUILHERME CASTRO BOULOS	1001472	0.145	5070	ERIKA SANTOS SILVA	256903	0.172
0.837	1520	SIMONE APARECIDA CURRALADAS DOS SANTOS	97730	0.446	4510	VITOR LIPPI	106661	0.458
0.826	4433	KIM PATROCA KATAGUIRI	295460	0.135	5070	ERIKA SANTOS SILVA	256903	0.172
0.822	1818	MARIA OSMARINA MARINA DA SILVA VAZ DE LIMA	237526	0.171	4433	KIM PATROCA KATAGUIRI	295460	0.135
0.817	4433	KIM PATROCA KATAGUIRI	295460	0.135	5010	GUILHERME CASTRO BOULOS	1001472	0.145
0.812	3030	ADRIANA MIGUEL VENTURA	109474	0.220	4433	KIM PATROCA KATAGUIRI	295460	0.135

Table 1: 10 Most Similar Pairs among Elected Candidates (SP 2022)

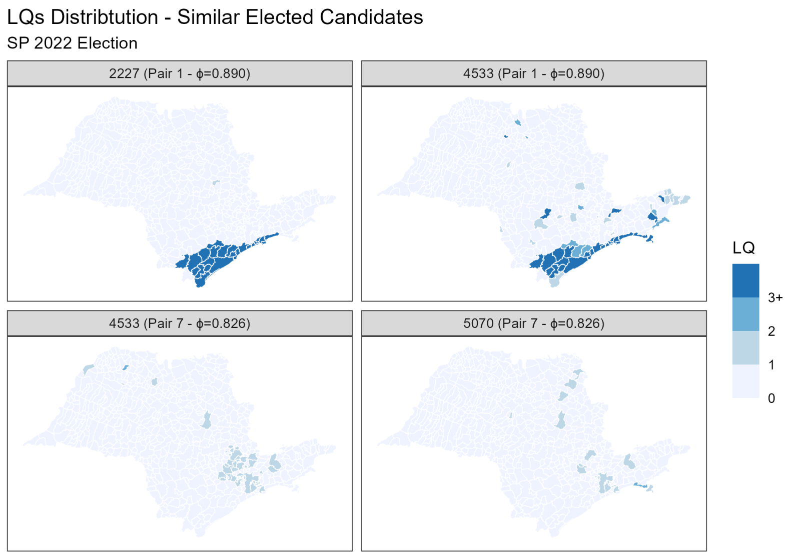

Figure 5: Candidate Similarity — Examples

Cross-sectional fit. The cross-sectional regression of observed LQ on candidate-space and municipality-space density produces strong results across all 182 state-elections in the sample. Figure 6 displays the R^2 values as a heatmap by state and election. For the full candidate sample, the median R^2 across state-elections is approximately 0.50, with most values falling between 0.35 and 0.65. For competitive candidates, the median R^2 rises to approximately 0.58, reflecting the removal of marginal candidates whose sparse vote vectors add noise. The largest states — SP, MG, RJ — tend to have lower R^2 values, consistent with the expectation that larger, more fragmented candidate fields are harder to predict. Smaller states, particularly in the North and Northeast, show R^2 values above 0.60.

Figure 7 further complements the hetamaps, showing the distribution of R^2 values pooled across states for each election year.[6] There is a downward trend (i.e., a shift to the left) from 1998 to 2022, both in the full and the competitive-candidate samples. This decline in cross-sectional fit is consistent with the temporal shift in the relative weight of the two density components documented below. As the electoral market moves from place-based to candidate-based segmentation — with municipality-space density losing explanatory power and candidate-space density gaining it — the overall structure becomes harder to predict from a single cross-sectional regression, even as the candidate-space component becomes individually more informative.

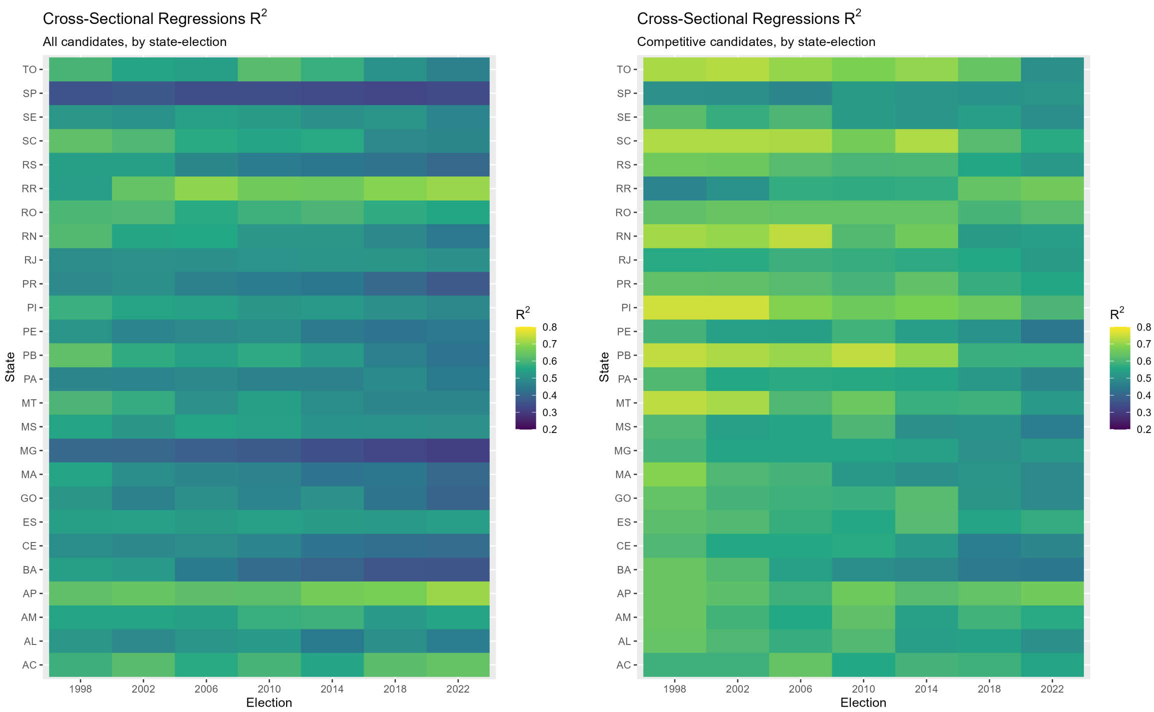

Figure 6: Cross-Sectional Regressions — R^2

The relative weight of candidate-space and municipality-space density. Figure 8 plots the coefficients on candidate-space density (β_C) against municipality-space density (β_L) for each state-election. A striking pattern emerges: states separate into distinct clusters. In most states, municipality-space density carries a coefficient near or above 1, indicating that knowing which municipalities are politically similar is highly informative. Candidate-space

[6]See complete summary table by election at table B.2 in the appendix.

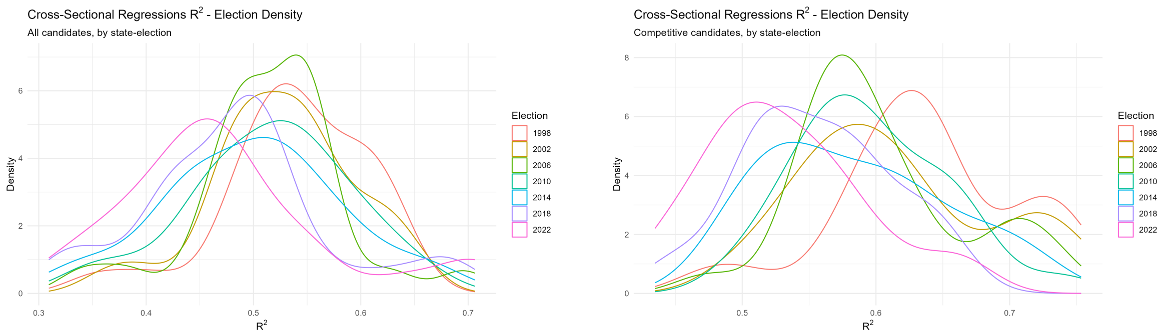

Figure 7: Cross-Sectional Regressions — R^2 Density

density is more variable: in states like RJ, AP, and RR, it dominates (coefficients above 0.8), while in many Northeastern states, it is close to zero and municipality-space density does virtually all the explanatory work. This variation suggests that the latent political structure is organized differently across states — in some states, the electoral market is segmented primarily by candidate type, while in others it is segmented primarily by place.

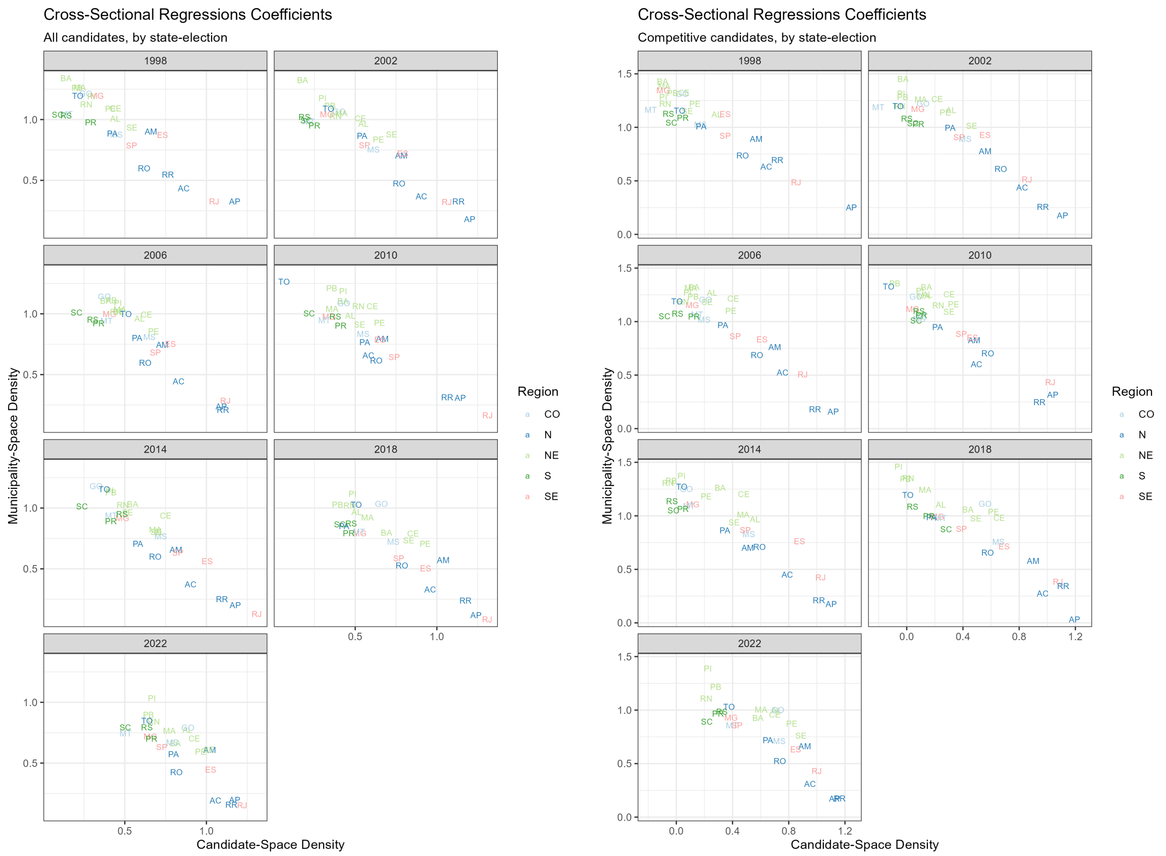

Figure 8: Cross-Sectional Regressions — Coefficients

The figure also reveals a notable temporal pattern: in most states, the coefficient on municipality-space density declines over time while the coefficient on candidate-space density increases. That is, knowing which municipalities are politically similar has become less informative for predicting vote intensity over the period, while knowing which candidates share similar profiles has become more informative. This shift is consistent with the hypothesis that the spatial structure of political information has flattened — as digital media and nationalized political issues erode the distinctiveness of place-based political environments, the electoral market becomes increasingly segmented by candidate type rather than by municipal characteristics.

4.2 Convergence

Consecutive election pairs. Table 2 presents the convergence regression results for all six consecutive election pairs (1998–2002 through 2018–2022). In every pair, the median coefficient on the residual across 26 states is negative and substantial (ranging from −0.515 to −0.583), and the share of states with a significant negative coefficient is 100%. The median R^2 of the convergence regression is approximately 0.15, meaning the base-year residual alone explains about 15% of the subsequent change in vote intensity. These results hold with negligible qualitative differences when restricted to competitive candidates (see table B.3 in the appendix).

Base year	End year	Sample	Median coef.	Mean coef.	Share neg. & sig.	Median R^2	N states
1998	2002	All	-0.583	-0.587	1	0.160	26
2002	2006	All	-0.557	-0.540	1	0.154	26
2006	2010	All	-0.545	-0.532	1	0.146	26
2010	2014	All	-0.515	-0.521	1	0.151	26
2014	2018	All	-0.556	-0.547	1	0.155	26
2018	2022	All	-0.551	-0.541	1	0.174	26

Table 2: Convergence regressions: consecutive election pairs

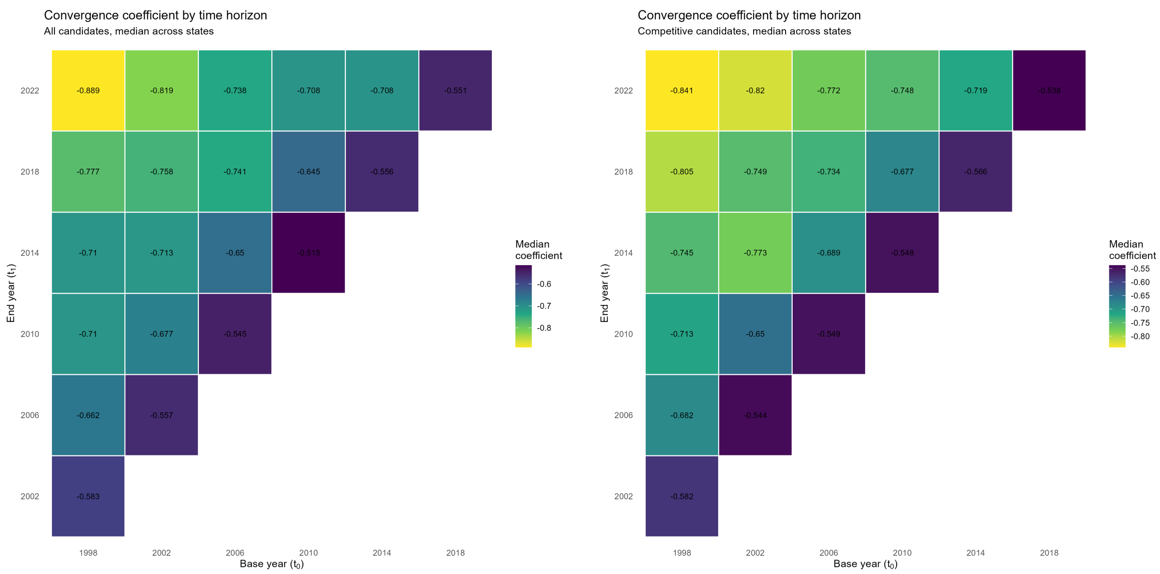

Figure 9: Convergence Regressions — Coefficients by Time Horizon

Time horizon. Figure 9 displays the median convergence coefficient for all 21 possible election-pair combinations, arranged by base year and end year. Every cell is negative.

Crucially, the coefficient becomes more negative as the time horizon lengthens: moving down any column (fixing the base year, extending the horizon), the convergence strengthens monotonically. Starting from 1998, the coefficient goes from -0.583 (4-year horizon to 2002) to -0.889 (24-year horizon to 2022). This pattern confirms the theoretical prediction: *the gap between observed and implied comparative advantage closes over time, and more time means more closing.*

The horizon scatterplot (Figure 10) summarizes this relationship: each point is a base-year/end-year pair, plotted against the time gap. The linear fit shows a clear downward slope, with the coefficient becoming approximately 0.06 units more negative for each additional 4-year electoral cycle.

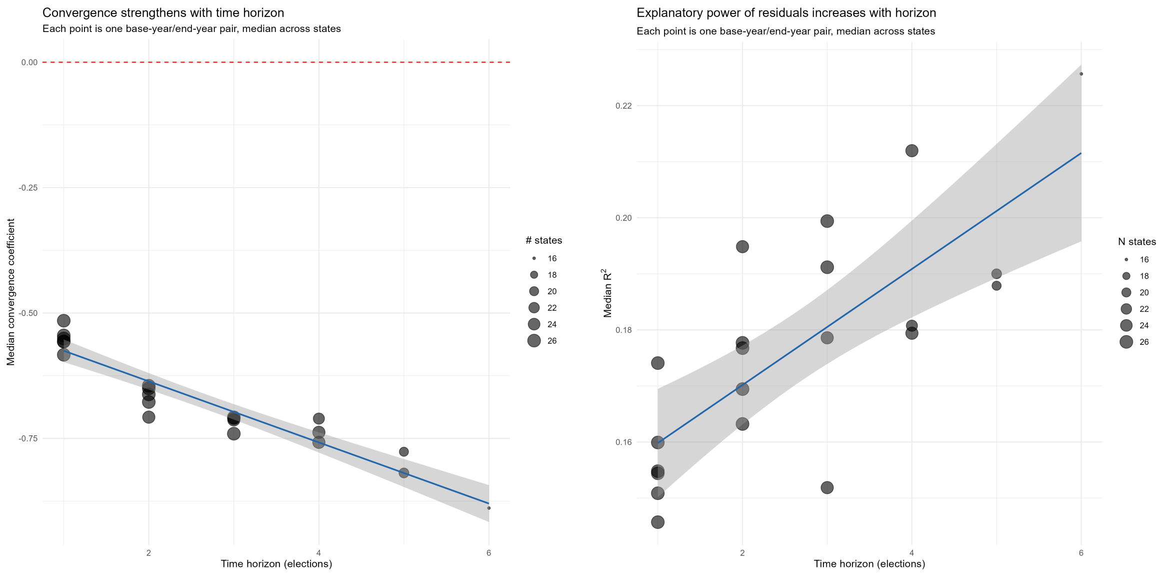

Figure 10: Convergence Regressions — Coefficients and R^2 by Time Horizon

This finding rules out simple mean reversion as an explanation. Mean reversion would produce the strongest convergence at short horizons, with the effect fading as the initial "shock" dissipates. The opposite pattern — stronger convergence at longer horizons — implies that the residuals capture slow-moving structural forces rather than transitory noise. The latent structure of candidate-municipality matching, as recovered by the implied comparative advantage framework, predicts the direction of electoral geography change over horizons of up to 24 years (or 6 electoral cycles).

Cross-state variation. Figure 11 shows the convergence coefficient by state for the 2018–2022 pair. All 26 states have negative coefficients; all are significant at the 5% level[7]. The strongest convergence is in the smaller Northern states (MA, RR, PI), where coefficients exceed -0.65; the weakest is in ES, RS, and SC. The full state × election-pair heatmap (Figure 13 in the appendix) confirms that convergence is a pervasive feature of Brazilian electoral geography: across more than 500 state-pair regressions, the coefficient is negative in the vast majority of cases.

[7]Clustered-standard errors at the candidate level.

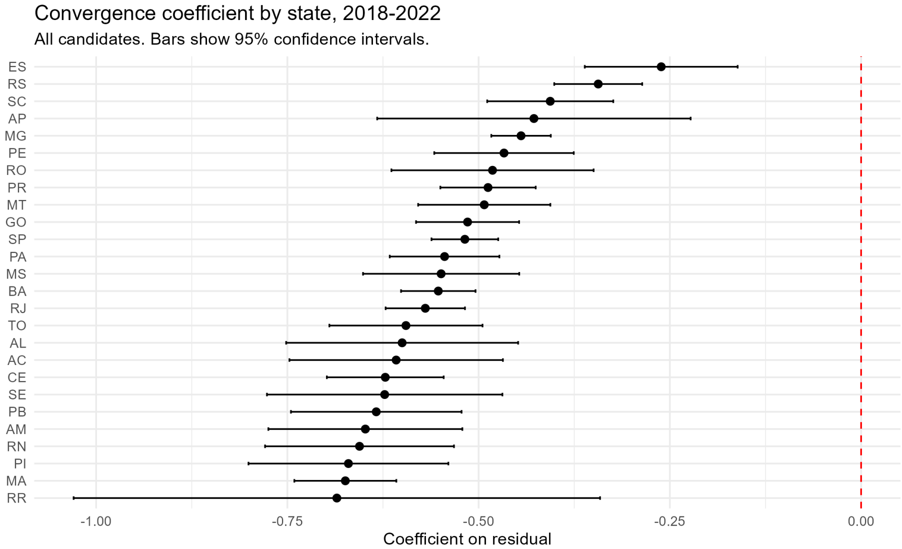

Figure 11: Convergence Regressions — Coefficients by State (2018-2022)

The levels specification, while necessary to accommodate zero-valued observations (i.e., the fact that many candidates do not receive votes in all municipalities in their states), may understate the model's fit relative to a log specification. As a robustness check, I restrict to candidate-municipality pairs with positive LQ in both periods for the convergence analysis in logs, which yields qualitatively identical results.

4.3 Extensive Margin

The extensive margin analysis tests whether the implied measures predict the *appearance* of candidates in municipalities where they were previously absent, and the *disappearance* of candidates from municipalities where they were previously present.

Table 3 reports the results for all six consecutive election pairs. For appearances, the median coefficient on the pseudo-residual is negative in every period, as expected: candidate-municipality pairs that are "unexpectedly absent" (the model predicts presence but the candidate receives no votes) are more likely to see the candidate appear in the subsequent election. The share of states with the correct sign ranges from 85% to 96%. For disappearances, the median coefficient is positive in every period: "unexpectedly present" pairs are more likely to see the candidate disappear in the municipality. The correct-sign share ranges from 96% to 100%. The appearance and disappearance rates provide context for these results. In a typical consecutive election pair, about 35–45% of initially absent candidate-municipality pairs see the candidate appear, while 15–28% of initially present pairs see the candidate disappear.

Base year	End year	Margin	Median coef.	Share correct sign	Median Appear Rate	Median Disappear Rate	N states
1998	2002	Appearance	-0.605	0.885	0.449	-	26
1998	2002	Disappearance	0.155	1.000	-	0.149	26
2002	2006	Appearance	-0.782	0.962	0.447	-	26
2002	2006	Disappearance	0.223	1.000	-	0.152	26
2006	2010	Appearance	-0.621	0.923	0.361	-	26
2006	2010	Disappearance	0.238	1.000	-	0.164	26
2010	2014	Appearance	-0.667	0.846	0.396	-	26
2010	2014	Disappearance	0.249	0.962	-	0.165	26
2014	2018	Appearance	-0.686	0.846	0.446	-	26
2014	2018	Disappearance	0.186	1.000	-	0.178	26
2018	2022	Appearance	-0.492	0.846	0.332	-	26
2018	2022	Disappearance	0.286	0.962	-	0.284	26

Table 3: Extensive margin: appearances and disappearances

5 Discussion

5.1 What the Framework Reveals

The results demonstrate that a substantial latent structure underlies the spatial distribution of votes in Brazilian federal deputy elections. This structure is recoverable from the cross-sectional pattern of vote intensities using the implied comparative advantage methodology, and it is remarkably stable: residuals from the 1998 cross-section predict vote geography changes 24 years (or 6 elections) later. Four features of this structure deserve emphasis.

First, the latent political landscape has spatial organization that transcends physical proximity. Municipality similarity decays with distance but captures variation that geographic distance alone cannot explain. Municipalities in different geographic regions of a state can be politically similar because they share, for instance, socioeconomic profiles, information environments, or party structures — features that produce similar patterns of candidate support without requiring geographic contiguity. Importantly, the framework recovers this political similarity without requiring the researcher to specify which contextual or compositional factors drive it.

Second, the organization of political competition differs systematically across states. In some states, knowing which other candidates resemble a given candidate is the primary source of explanatory power (candidate-space density dominates); in others, knowing which municipalities are similar matters more (municipality-space density dominates). This variation likely reflects differences in how electoral markets are segmented — by candidate profile, by territorial political machines, or by some combination of both. States where candidate-space density dominates may have electoral competition organized around distinct candidate types competing for overlapping geographic constituencies; states where municipality-space density dominates may have competition organized around territorial niches where local conditions are the primary determinant of which candidates thrive.

Third, the relative weight of these two sources of information shifts over time. In most states, the coefficient on municipality-space density has declined across elections while the coefficient on candidate-space density has increased. This means that knowing which municipalities are politically similar has become less informative for predicting vote intensity, while knowing which candidates share similar profiles has become more informative. This temporal shift is consistent with the hypothesis that the spatial structure of political information has flattened over the period — as digital media expand reach beyond geographically bounded networks and as party brands become more nationally legible, the electoral market becomes increasingly segmented by candidate type rather than by municipal characteristics. This interpretation resonates with Guarnieri and da Silva (2022) model of vote dispersion through information diffusion: as the channels of diffusion change, so does the structure of the latent space that the framework recovers.

Fourth, the convergence of observed toward implied comparative advantage strengthens over time. This is consistent with the interpretation that the implied measure captures durable features of the candidate-municipality match — features that exert a gradual gravitational pull on electoral geography despite the turbulence of individual elections, partisan realignments, and changing candidate fields. The fact that convergence operates over horizons of up to 24 years — a period spanning the PT's rise (and fall), Lula's presidency, Dilma's impeachment, and the Bolsonaro realignment — suggests that beneath the partisan turbulence, there exists a stable latent structure of political demand and supply that candidates gradually converge toward.

5.2 Scope Conditions and Limitations

The framework is best suited to electoral systems where individual candidates compete for votes across multiple geographic units within a district — open-list proportional representation being the canonical case. In closed-list systems where voters choose parties rather than individuals, the candidate $\times$ municipality matrix would collapse to a party $\times$ municipality matrix, which might still yield meaningful results but with substantially less variation on the candidate dimension.

The convergence analysis is necessarily restricted to candidates who run in multiple elections, which introduces a classic strategic selection dynamic. Candidates who exit politics after a single cycle — particularly those who suffer severe electoral defeats — are systematically excluded from the multi-election panel. While this omission likely understates the baseline convergence effect by purging the most severely misaligned candidate-municipality pairs from the data, it also introduces an important nuance when interpreting the incumbency dynamics explored in Appendix A. Because non-incumbents face immense financial and institutional hurdles to mount a subsequent campaign, challengers who successfully run across multiple cycles are disproportionately "strong types" with highly viable campaigns. The rapid upward mobility of these surviving challengers toward their latent potential may look like accelerated structural convergence. Conversely, safe incumbents possess the legislative resources and pork-barrel funding to artificially sustain idiosyncratic, non-structural regional bases over several cycles. Therefore, what appears to be slower structural convergence among elected officials may actually reflect a combination of deliberate, strategic resource targeting by officeholders and survivorship bias among highly competitive challengers.

Finally, while the baseline cross-sectional framework remains agnostic about the specific operational channels of electoral geography, the empirical extensions developed in Appendix A explicitly adjudicate between several of these competing mechanisms. By leveraging localized 3G technological rollouts and tracking legislative resource allocations, the analysis demonstrates that convergence is not a purely mechanical data artifact, nor is it driven by simple, supply-side incumbent expansion. Instead, the evidence points toward a structural model where technological shocks actively reshape the underlying equilibrium landscape, and institutional resources act as strategic frictions that officeholders use to protect idiosyncratic bases. Nevertheless, because the current empirical strategy operates at the macro-municipal level, it cannot fully isolate individual-level voter information diffusion from localized political peer effects (such as candidates actively mimicking the geographic footprints of successful co-partisans). Disentangling these fine-grained behavioral interactions from broader structural economic realignments remains a highly promising direction for future research

5.3 Implications

The implied comparative advantage framework produces, for every candidate-municipality pair in an electoral district, a decomposition of vote intensity into a structural component (what the latent political landscape predicts) and an idiosyncratic component (what is specific to this candidate in this place). This dyadic decomposition opens several avenues for future research.

In the study of distributive politics, the residual map provides a measure of *unrealized electoral potential* — municipalities where a deputy's profile is well-matched but their actual vote share is low. Whether and how deputies respond to this latent potential through targeted spending and constituency service is an empirical question that the framework makes tractable. The connection between electoral geography and parliamentary behavior could be revisited at the municipality level using the structural and idiosyncratic components of each deputy's vote geography.

In the study of party strategy, the candidate similarity matrix speaks directly to questions about the regionalization of candidate lists. For instance, Silotto (2019) has shown that parties organize their slates to minimize intra-party geographic competition. The candidate similarity matrix recovers this structure from the vote data itself — without requiring information about candidates' career trajectories or party nominations — and could be used to measure the degree of geographic coordination within and across party lists.

In the study of electoral dynamics, the convergence result suggests that a component of the long-term evolution of electoral geography is already implied in today's patterns of voting. The municipality similarity matrix provides a data-driven map of how the political landscape is organized — which places share a common political character, and which are politically distinct — without requiring pre-specified demographic or ideological categories. Far from being a static baseline, the framework — paired with the empirical extensions in the appendix — demonstrates that this latent landscape represents a moving equilibrium continually negotiated between shifting informational boundaries and institutional protections. Tracking the temporal evolution of these similarity matrices over time therefore provides a precise, measurable tool to observe exactly how, where, and how fast systemic shocks and deep partisan realignments reshape the geographic foundations of political competition.

6 Conclusion

This paper introduces the implied comparative advantage framework to the study of electoral geography, adapting Hausmann et al. (2022)'s methodology from economic geography to recover the latent structure of candidate-municipality matching from aggregate vote data. Using data from all Brazilian states across seven federal deputy elections (1998–2022), I show that implied vote intensity measures explain a substantial share of observed vote intensity cross-sectionally, that the residuals from this relationship predict future vote geography changes over horizons of up to 24 years, and that this convergence operates at both the intensive and extensive margins.

The framework produces objects that are new to electoral geography: a measure of implied vote intensity for every candidate-municipality pair, similarity matrices that reveal the latent organization of political competition across municipalities and across candidates, and dyadic residuals that decompose vote geography into structural and idiosyncratic components. Beyond these measurement contributions, the empirical findings carry substantive implications. The existence of a stable, recoverable latent structure underlying Brazilian electoral geography suggests that the spatial distribution of votes is not merely the product of individual campaign strategies and electoral contingencies. It reflects a deeper pattern of political demand and supply — a match between what municipalities need and what candidates offer — that asserts itself over decades despite the turbulence of Brazilian politics. And the temporal shift from municipality-based to candidate-based predictability suggests that the organization of this latent structure is itself evolving, in ways that are consistent with the transformation of political information environments over the past quarter-century.

Bibliography

Agnew, J. (1996). Mapping politics: how context counts in electoral geography. *Political Geography*, 15(2):129–146.

Ames, B. (1995). Electoral Strategy under Open-List Proportional Representation. *American Journal of Political Science*, 39(2):406–433.

Ames, B. (2001). *The deadlock of democracy in Brazil*. Interests, identities, and institutions in comparative politics. University of Michigan Press, Ann Arbor.

Avelino, G., Biderman, C., and Silva, G. P. d. (2011). A Concentração eleitoral nas eleições paulistas: medidas e aplicações. *Dados*, 54:319–347.

Avelino, G., Biderman, C., and Silva, G. P. d. (2016). A Concentração Eleitoral no Brasil (1994-2014). *Dados*, 59:1091–1125.

Azevedo, D. A. d. (2023). The need for electoral geography: the possibilities in the field. *GEOUSP*, 27:e.

Bräuninger, T., Däubler, T., and Pilet, J.-B. (2024). Candidate visibility, voter knowledge, and the incumbency advantage in preferential-list PR. *Party Politics*, 30(1):61–72.

Carey, J. M. and Shugart, M. S. (1995). Incentives to cultivate a personal vote: A rank ordering of electoral formulas. *Electoral Studies*, 14(4):417–439.

Carvalho, N. R. d. (2003). *E no início eram as bases: geografia política do voto e comportamento legislativo no Brasil*. Editora Revan. Google-Books-ID: Cn0PAAAAYAAJ.

Cheibub, J. A. and Nalepa, M. (2020). Revisiting electoral personalism. *Journal of Theoretical Politics*, 32(1):3–10.

Clinton, J., Jackman, S., and Rivers, D. (2004). The Statistical Analysis of Roll Call Data. *American Political Science Review*, 98(2):355–370.

Cox, G. W. (1990). Centripetal and Centrifugal Incentives in Electoral Systems. *American Journal of Political Science*, 34(4):903–935.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper. Google-Books-ID: kLEGAAAAMAAJ.

Duda, R. O. (2001). *Pattern classification*. Wiley-Interscience publication. Wiley, New York, N.Y, second edition. edition.

Enelow, J. M. and Hinich, M. J. (1989). A General Probabilistic Spatial Theory of Elections. *Public Choice*, 61(2):101–113.

Forest, B. (2018). Electoral geography: From mapping votes to representing power. *Geography Compass*, 12(1):e12352. _eprint: https://compass.onlinelibrary.wiley.com/doi/pdf/10.1111/gec3.12352.

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

Guarnieri, F. and da Silva, G. P. (2022). A spatial interaction model of vote dispersion. *Political Geography*, 98:102709.

Hausmann, R., Stock, D. P., and Yıldırım, M. A. (2022). Implied comparative advantage. *Research Policy*, 51(8):104143.

Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. (2007). The Product Space Conditions the Development of Nations. *Science*, 317(5837):482–487.

Jankowski, M. (2016). Voting for locals: Voters' information processing strategies in open-list PR systems. *Electoral Studies*, 43:72–84.

Key, V. O. (1949). *Southern Politics in State and Nation*. A. A. Knopf. Google-Books-ID: PlKJ0QEACAAJ.

King, G. (1996). Why context should not count. *Political Geography*, 15(2):159–164.

Lupia, A. and McCubbins, M. D. (1998). *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge University Press. Google-Books-ID:

2Vv6BhLC6HUC.

Mayhew, D. R. (1974). *Congress: the electoral connection.* Number 26 in Yale studies in political science. Yale University Press, New Haven.

Myerson, R. B. (1993). Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems. *The American Political Science Review*, 87(4):856–869.

O'Donoghue, D. and Gleave, B. (2004). A Note on Methods for Measuring Industrial Agglomeration. *Regional Studies*, 38(4):419–427.

Panzera, D., Cartone, A., and Postiglione, P. (2022). New evidence on measuring the geographical concentration of economic activities. *Papers in Regional Science*, 101(1):59–79.

Poole, K. T. and Rosenthal, H. (1985). A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science*, 29(2):357–384.

Put, G.-J., von Schoultz, A., and Isotalo, V. (2020). Fighting over friends and neighbors: The effect of inter-candidate geographic distance on intra-party competition. *Political Geography*, 81:102219.

Rocha, C. V. (2021). "Conexão Eleitoral", geografia do voto e produção legislativa: um estudo de caso. *Cadernos Metrópole*, 23(51):581–604.

Shugart, M. S., Valdini, M. E., and 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. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0092-5853.2005.00133.x.

Silotto, G. (2019). A relevância regional nas estratégias partidárias: evidências das listas de candidatos de São Paulo. *Revista de Sociologia e Política*, 27:e003.

Silva, G. P. d. and Silotto, G. (2018). Preparing the Terrain: Conditioning Factors for the Regionalization of the Vote for Federal Deputy in São Paulo. *Brazilian Political Science Review*, 12.

Stokes, D. E. (1963). Spatial Models of Party Competition. *American Political Science Review*, 57(2):368–377.

A Exploring the Mechanisms Behind Convergence

The main results establish that convergence (i.e., the tendency for observed vote geography to move toward the structural prediction of the implied comparative advantage model) is a robust feature of Brazilian electoral geography. It holds across all states, in every election pair, and strengthens monotonically with the time horizon. This appendix draws on the microfoundations developed in Section 2 to assess what kind of process generates this convergence, and what it tells us about the nature of the latent structure the framework recovers.

Deriving predictions

The microfoundations suggest that the latent match structure between candidates and municipalities is shaped by the interaction of voter preferences, the spatial structure of political information, and candidates' strategic positioning (including the coordination of party lists). Deviations from the structural prediction — the residuals — reflect frictions: information barriers that prevent voters from discovering well-matched candidates, coordination failures among co-partisans, campaign resource constraints, and incumbents' ability to sustain idiosyncratic vote geographies through targeted spending. Convergence reflects the gradual resolution of these frictions as actors learn, adapt, and adjust over successive elections.

I consider two comparative statics that follow from this framework:

Incumbency. If convergence is partly driven by elected deputies' strategic investments in structurally promising municipalities (e.g., using budgetary amendments, constituency service, and other distributive instruments to build support where their profile is well-matched) then incumbents should converge faster than non-elected candidates. However, a countervailing force operates in the opposite direction: the same resources that could accelerate convergence can also be used to *sustain* deviations from the structural prediction, defending idiosyncratic strongholds that would otherwise erode. The net effect is an empirical question.

Information environment. The implied comparative advantage at time t_0 captures the latent structure as it stands in that election (i.e., a structure shaped, in part, by the prevailing information environment). If the information environment changes substantially between t_0 and t_1, as when municipalities gain access to mobile broadband, two effects operate. On one hand, reduced information barriers could accelerate convergence by helping voters discover well-matched candidates they previously could not evaluate. On the other hand, the information shock may alter the latent structure itself: candidates who lacked comparative advantage under the old, spatially structured information regime may find it easier and less costly to reach voters in new municipalities once digital channels are available, effectively reshaping the match between candidate profiles and municipal electorates. Under this second interpretation, the t_0 structural prediction becomes a less reliable guide to the direction of change, and convergence toward it weakens (not because convergence as a process has slowed, but because the target has shifted).

Empirical strategy

For the incumbency test, I interact the cross-sectional residual with an indicator for whether the candidate was elected in t_0:

$$\frac{R_{cl,t_1} - R_{cl,t_0}}{(t_1 - t_0)/4} = \alpha + \beta_1 \hat{\varepsilon}_{cl,t_0} + \beta_2(\hat{\varepsilon}_{cl,t_0} \times \text{Elected}_{c,t_0}) + \gamma \cdot \text{Elected}_{c,t_0} + e_{cl} \tag{10}$$

The specification is estimated separately for each state-election pair, and I report the distribution of β_2 across state-election pairs. I also report specifications controlling for the initial LQ level and its interaction with incumbency.

For the information environment test, I exploit variation in the timing of 3G mobile broadband deployment across Brazilian municipalities, using data from ANATEL (Brazil's National Telecommunications Agency). I discretize the date of 3G arrival to match election years and define an indicator $\text{Gained3G}_{l,t}$ equal to one if municipality l gained 3G coverage between elections t_0 and t_1. I restrict the sample to municipalities that did not have 3G at t_0 (to compare municipalities that gained coverage during the period to those that remained unconnected) and estimate:

$$\frac{R_{cl,t_1} - R_{cl,t_0}}{(t_1 - t_0)/4} = \alpha_s + \beta_1 \hat{\varepsilon}_{cl,t_0} + \beta_2(\hat{\varepsilon}_{cl,t_0} \times \text{Gained3G}_{l,t}) + \gamma \cdot \text{Gained3G}_{l,t} + e_{cl} \tag{11}$$

where α_s are state fixed effects and standard errors are clustered at the candidate level. The residuals are standardized within each state-election (divided by the within-state standard deviation) to ensure comparability when pooling across states.

Results

Incumbency. Table A.1 reports the distribution of the interaction coefficient β_2 across consecutive election pairs. The median coefficient is positive in most periods, and the share of states with a significant positive interaction consistently exceeds the share with a significant negative interaction. This means that elected candidates converge *slower* than non-elected candidates toward the structural prediction. The result is consistent across both the full and competitive candidate samples, and robust to controlling for initial LQ levels.

Albeit not strong (as most estimated interactions are not statistically significant), this finding suggests that the resources of office (such as budgetary amendments, constituency service, the parliamentary allowance) allow elected deputies to sustain vote geographies that deviate from the latent match structure. Rather than using distributive instruments to expand into structurally promising municipalities, deputies appear to reinforce their existing coalitions, maintaining idiosyncratic strongholds that would otherwise erode. This is consistent with the emphasis in the Brazilian distributive politics literature on the *maintenance* rather than *expansion* logic of legislative targeting (Ames, 2001; Carvalho, 2003). From the perspective of the convergence process, supply-side resources act as a friction that slows adjustment rather than accelerating it.

Base year	End year	Sample	Median coef.	Mean coef.	Share pos. & sig.	Share neg. & sig.	Median R^2	N states
1998	2002	All	0.029	0.017	0.077	0.077	0.175	26
2002	2006	All	-0.012	0.029	0.038	0.000	0.156	26
2006	2010	All	-0.029	-0.010	0.038	0.038	0.154	26
2010	2014	All	0.025	0.027	0.080	0.040	0.161	25
2014	2018	All	0.038	0.061	0.115	0.000	0.166	26
2018	2022	All	0.040	0.069	0.154	0.000	0.179	26
1998	2002	Competitive	0.095	0.096	0.240	0.080	0.154	25
2002	2006	Competitive	0.088	0.039	0.130	0.087	0.149	23
2006	2010	Competitive	0.048	0.086	0.115	0.077	0.167	26
2010	2014	Competitive	0.082	0.094	0.208	0.000	0.162	24
2014	2018	Competitive	0.106	0.108	0.269	0.038	0.165	26
2018	2022	Competitive	0.124	0.133	0.269	0.000	0.174	26

Table A.1: Convergence and Incumbency: consecutive election pairs

Information environment. Table A.2 reports the interaction between the convergence residual and the 3G indicator for election pairs where meaningful variation in 3G deployment exists (2010–2014, 2014–2018). The interaction coefficient is positive and significant: municipalities that gained 3G access during the inter-election period show weaker convergence toward the t_0 structural prediction.

This result is inconsistent with a simple information-diffusion account in which digital connectivity helps voters discover candidates who were already well-matched under the existing structure. Instead, it is consistent with the interpretation that the expansion of digital connectivity reduces the cost for candidates to explore new electoral ground (e.g., reaching municipalities where they lacked comparative advantage under the old, spatially structured information regime). The arrival of 3G effectively reshapes the match between candidate profiles and municipal electorates, making the pre-shock structural prediction a less reliable guide to the direction of change.

Importantly, this result does not undermine the implied comparative advantage framework. The framework does not require the latent structure to be permanent; only that it evolves slowly enough for a cross-sectional snapshot to remain informative over multiple election cycles. The time-horizon results in the main paper provide strong evidence for this: the 1998 cross-section predicts vote geography changes through 2022 more powerfully than over a single cycle, despite 24 years of political, economic, and technological transformation. The 3G result provides a complementary piece of evidence from the other direction: in municipalities where the information environment undergoes rapid change, convergence toward the *pre-existing* prediction weakens. This is what the framework should predict (i.e., the cross-sectional snapshot is less informative where the underlying structure is evolving faster) and it confirms that the convergence pattern documented in the main results reflects genuine structural content rather than mechanical mean reversion.

	2010–2014		2014–2018	
	All	Competitive	All	Competitive
Model:	(1)	(2)	(3)	(4)
Variables				
Residual	-0.4971***	-0.5398***	-0.5506***	-0.6268***
	(0.0132)	(0.0152)	(0.0219)	(0.0262)
Gained 3G	-0.0448***	-0.0299**	-0.0258	-0.0222
	(0.0141)	(0.0149)	(0.0220)	(0.0246)
Residual × Gained 3G	0.0785***	0.0926***	0.0648***	0.1059***
	(0.0124)	(0.0144)	(0.0227)	(0.0271)
Fixed-effects				
State	Yes	Yes	Yes	Yes
Fit statistics				
Observations	111,586	101,261	35,813	31,612
R^2	0.16681	0.16672	0.18022	0.16891

Clustered (by Candidate) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table A.2: Convergence and 3G Expansion: consecutive election pairs

Discussion

The two comparative statics point in a consistent direction: convergence is not primarily driven by identifiable supply-side mechanisms. Incumbents' distributive resources slow it rather than accelerate it, and information-environment shocks disrupt the pre-existing prediction rather than reinforcing it. This suggests that the convergence documented in the main paper reflects the cumulative operation of electoral competition itself, such as the gradual sorting of candidates into niches where their profiles best match voter demand, operating through multiple channels (candidate entry and exit, party list adjustment, slow-moving changes in voter awareness) none of which individually dominates but which collectively produce the strong, universal pattern. The implied comparative advantage framework captures the slow-moving equilibrium toward which this process tends, while the residuals measure the frictions — some strategic, some informational — that sustain disequilibrium in any given election.

	Municipality i		Municiaplity i'	
ϕ	Name	Region	Name	Region
0.8210621	Santos	São Paulo	São Vicente	São Paulo
0.8171456	Americana	Campinas	Santa Bárbara d'Oeste	Campinas
0.8135797	Sorocaba	Sorocaba	Votorantim	Sorocaba
0.8132250	Jundiaí	Campinas	Várzea Paulista	Campinas
0.8053303	Jacareí	São José dos Campos	São José dos Campos	São José dos Campos
0.7970711	Embu das Artes	São Paulo	Taboão da Serra	São Paulo
0.7917733	Mirassol	São José do Rio Preto	São José do Rio Preto	São José do Rio Preto
0.7841941	Barueri	São Paulo	Carapicuíba	São Paulo
0.7835006	Cubatão	São Paulo	Santos	São Paulo
0.7827941	Hortolândia	Campinas	Campinas	Campinas

Table B.1: 10 Most Similar Pairs of Municipalities (SP 2022)

B Supplementary Tables and Figures

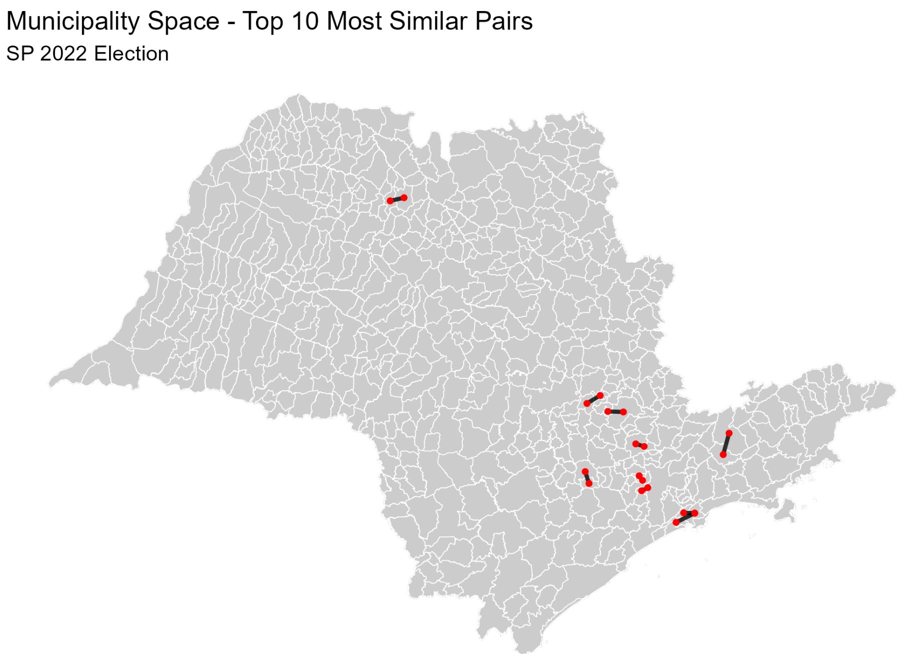

Figure 12: Spatial Distribution of the 10 Most Similar Pairs of Municipalities (SP 2022)

Year	Sample	Candidate Space		Municipality Space			
		Median coef.	Mean coef.	Median coef.	Mean coef.	Median R^2	Median R^2
1998	All	0.418	0.451	1.020	0.944	0.538	0.540
2002	All	0.537	0.552	0.958	0.847	0.532	0.527
2006	All	0.542	0.587	0.947	0.835	0.519	0.516
2010	All	0.524	0.553	0.936	0.866	0.520	0.515
2014	All	0.634	0.656	0.857	0.792	0.505	0.499
2018	All	0.676	0.718	0.793	0.724	0.487	0.484
2022	All	0.804	0.832	0.684	0.628	0.460	0.472
1998	Competitive	0.066	0.210	1.125	1.056	0.635	0.638
2002	Competitive	0.245	0.305	1.060	0.981	0.605	0.615
2006	Competitive	0.214	0.341	1.070	0.976	0.583	0.603
2010	Competitive	0.184	0.294	1.076	0.994	0.590	0.596
2014	Competitive	0.379	0.384	1.032	0.944	0.575	0.585
2018	Competitive	0.412	0.453	0.984	0.900	0.551	0.555
2022	Competitive	0.676	0.627	0.882	0.811	0.523	0.531

Table B.2: Cross-sectional Regressions: Coefficients and R^2 Summary

Base year	End year	Sample	Median coef.	Mean coef.	Share neg. & sig.	Median R^2	N states
1998	2002	Competitive	-0.582	-0.594	1	0.145	26
2002	2006	Competitive	-0.544	-0.545	1	0.144	26
2006	2010	Competitive	-0.549	-0.548	1	0.143	26
2010	2014	Competitive	-0.548	-0.534	1	0.148	26
2014	2018	Competitive	-0.566	-0.553	1	0.156	26
2018	2022	Competitive	-0.538	-0.547	1	0.162	26

Table B.3: Convergence regressions: consecutive election pairs

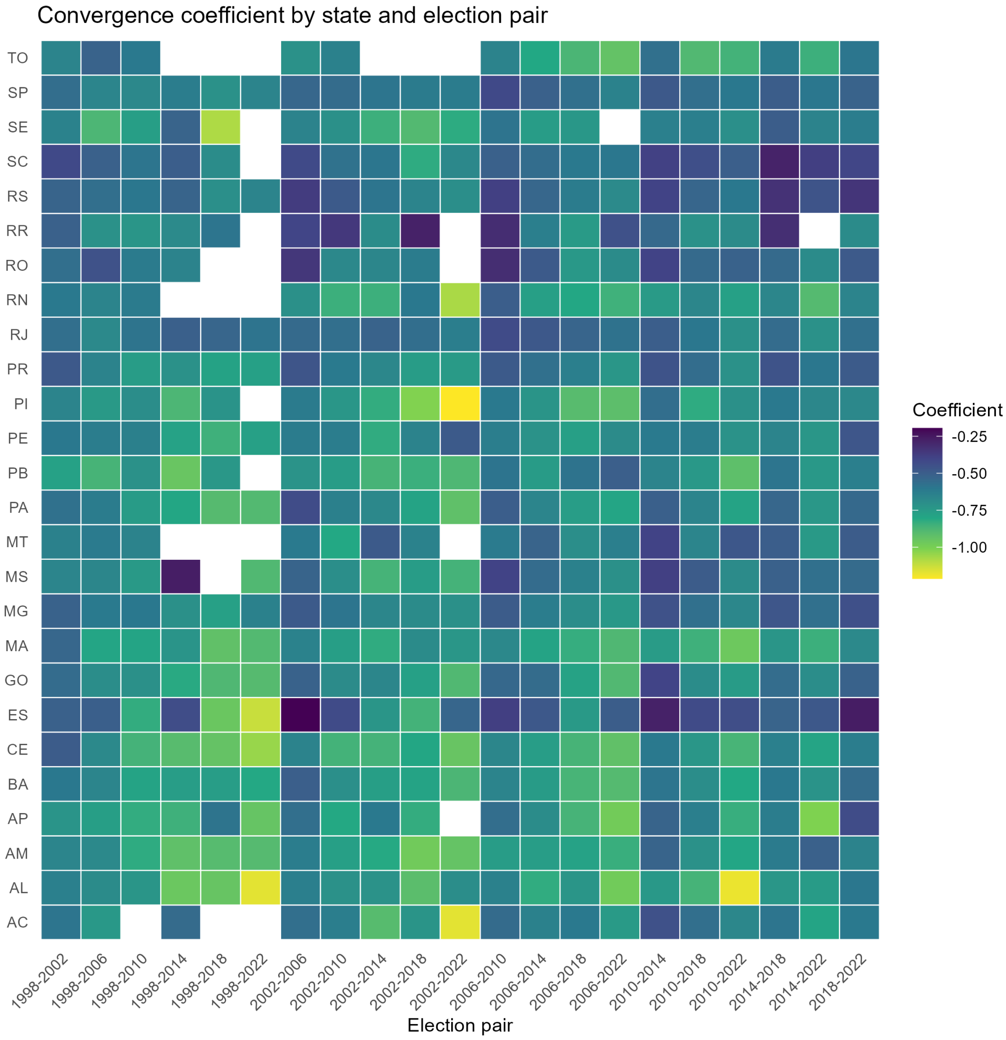

Figure 13: Convergence Regressions — Coefficients by State and Election Pair