

Ecological Inference for Electoral Analysis: A Computational Perspective on Human Decision-Making

J.M. Pavía¹, P. Pérez-Asurmendi², and R. de Andrés Calle³

¹ Department of Applied Economics, Universitat de València, Valencia, Spain
`Jose.M.Pavia@uv.es`

² PRESAD Research Group and IMUVA, Department of Economic Analysis,
Universidad Complutense de Madrid, Madrid, Spain
`patrizip@ucm.es`

³ BORDA and PRESAD Research Groups, Research Excellence Unit GECOS and
Multidisciplinary Institute of Enterprise (IME), Universidad de Salamanca,
Salamanca, Spain
`rocioac@usal.es`

Abstract. Ecological inference (EI) encompasses various methods to address the frequent lack of individual-level data in various areas, particularly in decision-making processes. Considering only aggregate information, these methods allow inferences to be drawn about individual behaviors and preferences. The difficulty in EI stems from the potential for ecological fallacy, where inferences drawn from aggregated data do not accurately reflect relationships at the individual level. Numerous statistical techniques and computational tools have been developed to address this challenge. This study focuses on comparing and contrasting several prominent R packages specifically designed for ecological inference: `lphom`, `eiPack`, `eiCircle`, `ecolRxC` and `eiopt2`. The comparison delves into the distinct computational approaches employed by each package, examines the underlying statistical models that form their foundation, and critically evaluates how each handles the inherent uncertainty associated with ecological inference. This includes considering how each package addresses the potential biases and limitations inherent in aggregated data. To provide a robust empirical assessment, this study leveraged paired survey data from the Spanish Center for Sociological Research (CIS) focusing on pre- and post-election studies related to the 2015 Spanish General Election.

Keywords: Electoral Decision-Making · Computational Modelling · Uncertainty.

1 Introduction

Intelligent Systems encounters a fundamental challenge in decision-making under conditions of uncertainty, particularly in the context of electoral and political analysis where comprehensive individual data is often unavailable. In such

scenarios, it becomes essential to infer individual behaviors from aggregated data, a task that introduces significant methodological complexities.

Ecological inference (EI) methods aim to reconstruct individual behaviors from aggregated data. Yet, they inherently face the risk of ecological fallacy, where patterns observed at the aggregate level may not accurately reflect individual-level decision-making [17]. Understanding voter transitions and political behavior through EI is crucial for interpreting how individuals make electoral choices in uncertain environments.

While classical decision models assume rational optimization, research in behavioral Economics and Psychology highlights deviations driven by heuristics, social influences, and cognitive biases. Voting behavior, in particular, is shaped by incomplete information and contextual factors, making it a complex process beyond pure rationality.

Over time, EI methods have evolved from regression-based models [5] and Bayesian approaches [7] to modern optimization techniques that enhance predictive accuracy and computational efficiency [18]. These advancements provide new perspectives on modeling voter behavior in uncertain contexts. Besides boosting computational efficiency, these methods also provide fresh perspectives on decision-making in uncertain situations.

This study evaluates and compares computational approaches to EI, focusing on R packages for voter transition analysis. It examines their ability to model uncertainty, integrate prior knowledge, and balance accuracy with efficiency. Considering the pre- and post-election studies related to the 2015 Spanish General Elections, provided by the CIS, which recorded declared voting intention and votes, respectively, this article offers a practical assessment of how different EI methods capture the complexity of electoral behavior. It is worth noting that the data used allow for comparing the different results of the different EI methods considered with actual data, given their panel data structure. Beyond voting research, these insights contribute to refining Intelligent Systems by incorporating uncertainty, heuristics, and adaptability into decision models.

The remainder of this paper is structured as follows. Section 2 reviews previous research and methodological advances in EI. Section 3 introduces the computational tools analyzed, covering their theoretical foundations and practical implementations. Section 4 presents the empirical application and initial hypotheses. Finally, Section 5 discusses the study’s findings and implications.

2 Background and Related Work

Ecological inference (EI) represents a significant methodological issue that entails drawing conclusions about individual behaviors based on data that is compiled at an aggregate level, particularly in social contexts. In this field, rational decision-making models are contrasted with observed decision-making processes in humans. The former assumes optimal choices based on complete information, whereas the latter often simplifying strategies and environmental information.

The application of EI in electoral studies and social preferences underscores its significance in understanding voter behavior and social dynamics.

In 1950, Robinson [17] challenged the use of ecological inference to address the ecological fallacy, which arises when conclusions drawn from aggregated data fail to accurately capture individual-level relationships. Since then, significant methodological advancements have emerged, beginning with Goodman’s regression-based approach (1953) [5] and followed by King’s seminal Bayesian framework (1997) [7], which fundamentally reshaped the field by introducing hierarchical Bayesian modeling to manage inherent uncertainties and opening a new era in the field.

Whilst traditional decision-making models under risk and uncertainty typically assume rationality and consistency (such as those based on the expected utility theory), extensive research in Psychology and behavioral Economics demonstrates systematic deviations from rational behavior in human decision-making [1]. The results highlight the need for improved EI techniques that are both mathematically sound and practically effective in capturing the complex human behaviors that rational models might miss.

EI has been extensively employed in electoral studies, primarily for the estimation of voter transition matrices derived from aggregate election data. Such estimations are vital for analyzing electoral shifts across different elections, deciphering patterns in voter behavior, assessing the impact of electoral campaigns, and providing detailed insights into political dynamics at various levels of aggregation (see [6] and [8]). Given the inherent complexity and uncertainty associated with aggregated electoral data, robust EI methods play a crucial role in supporting informed political decision-making and in providing reliable predictions of electoral outcomes.

Researchers have developed many new ways to make EI more accurate and dependable. These newer models are better at understanding uncertainty and how different election situations can vary [20]. Recent advancements incorporate computer-based methodologies, such as linear programming and integer optimization. These techniques are proficient in managing large electoral datasets and provide enhanced predictions regarding voter behavior changes (see [11] and [13]). Collectively, these advancements contribute to a more comprehensive understanding of voting patterns and electoral processes.

3 Ecological Inference Methods and Computational Tools

Given aggregate data structured in an $R \times C$ contingency table, where R and C are the number of voting alternatives in rows and columns, respectively, EI methods aim to estimate the unobserved internal cell counts or probabilities using the observed margins within subunits of the aggregate population. This estimation task typically involves solving underdetermined systems, where the number of unknown parameters exceeds the number of available aggregate observations, making explicit modeling assumptions essential.

Let Y_{kij} represent the number of individuals in group k (e.g., a geographic unit) belonging to category i and choosing issue j (e.g., a political party). The observed aggregate data consist of row totals $Y_{ki.} = \sum_{j=1}^C Y_{kij}$ and column totals $Y_{k..} = \sum_{i=1}^R Y_{kij}$ by group. The main objective is to estimate the unobserved proportions:

$$\theta_{ij} = \frac{Y_{.ij}}{Y_{..}}, \quad \text{for } i = 1, \dots, R, \quad j = 1, \dots, C.$$

The estimation of the θ_{kij} unknowns is also sometimes stated as a subsidiary objective. Since the number of equations provided by the observed marginal totals is insufficient to uniquely determine all θ_{ij} (and θ_{kij}), additional assumptions or statistical models are required to solve the EI problem.

Several computational methodologies have been developed to tackle the EI challenge systematically. This section introduces five prominent R packages specifically designed for ecological inference: **lphom**, **eiPack**, **eiCircle**, **ecolRxC**, and **eiopt2** (see Table 1). Each employs distinct methodological strategies to solve the underlying estimation problems:

lphom [12]: This package uses linear programming techniques to tackle constrained optimization problems effectively, assuming approximate homogeneity across k . This simplifies the complexity of the problem, allowing for the generation of deterministic point estimates of transition proportions. However, it is important to note that the reliance on mild homogeneity assumptions means that uncertainty is not explicitly modeled probabilistically. Instead, **lphom** addresses uncertainty through sensitivity analysis and bootstrap, which evaluates how changes in input parameters affect the outcomes. This approach helps users understand the robustness of the estimated proportions in the face of potential variations in data. The package provides different estimation functions, each tailored to specific modeling needs.

- **tslphom**: Implements a two-step estimation approach, refining initial global transition proportions θ_{ij} by generating local unit transition proportions θ_{kij} congruent with unit observed margins. This is particularly useful for combining aggregate and a priori information.
- **nsllphom**: Employs an iterative version of **tslphom**, refining transition proportions iteratively, offering greater flexibility. This is the workhorse algorithm of the **lphom**-approaches, generating robust estimates in general voting behavior contexts.
- **lclphom**: Uses a local iterative version of **nsllphom**, providing more flexibility to adapt the unit transition estimates to specific conditions. This is helpful when dealing with heterogeneous voting behaviors.
- **rsllphom**: Applies a re-scaling-based approach to implement a bottom-up method that improves the θ_{ij} by generating a θ_{kij} unit estimate at a time. This is computationally more intensive and recommended when unit estimates are of primary interest.

- eiCircles** [3]: This package estimates $R \times C$ vote transfer matrices from aggregate data using the model described in [3], an extension of the model proposed by Brown and Payne [2]. This package employs overdispersed-multinomial models, assuming that voters are grouped into units (e.g., concentric circles of family members, close friends, coworkers, and neighborhoods) that may influence each other through daily interactions [4]. The model allows for the incorporation of covariates to address unit heterogeneity and mitigate ecological bias.
- eiPack** [9]: This package employs hierarchical Bayesian models inspired by King’s ecological inference method [7], but addressing the complexities inherent in $R \times C$ contingency tables beyond the conventional 2×2 format [19]. It explicitly estimates transition probabilities through posterior probability distributions derived via Markov Chain Monte Carlo (MCMC) techniques. By incorporating both informative and non-informative priors, the method enhances flexibility, allowing for probabilistic interpretations of estimated outcomes. A notable advantage is its ability to quantify and communicate uncertainty explicitly, providing comprehensive insights into potential scenarios. This facilitates informed decision-making by enabling practitioners to assess risks and uncertainties inherent in ecological inference systematically.
- ecolRxC** [14]: This package addresses the ecological inference problem by utilizing latent structure theory, positing that voters’ decisions are influenced by unobserved individual latent variables as well as party popularity and positions on relevant issues [21]. Unlike other methods that predict expected behaviors, **ecolRxC** focuses on modeling how voters make choices based on their underlying preferences, providing a causal theory of political behavior to guide its algorithmic [15] approach. This makes it unique compared to other ecological inference techniques.
- eiOpt2** [10]: This package uses nonlinear quadratic optimization techniques, minimizing Euclidean distances between standardized row proportions for each unit and overall standardized row proportions. Optimization is performed under constraints using algorithms from the R packages **alabama** and **quadprog**.

Note that other packages available for EI, such as **ei** and **eco**, have not been considered because they are limited to 2×2 tables.

Given the methodological challenges, establishing an objective accuracy measure for EI algorithms is crucial. Existing evaluation strategies have limitations, such as biases and context dependence [8,16]. To address these, the Error Index [11] is introduced as an intuitive metric to quantify the discrepancy between estimated and actual transition matrices.

Let Y_{ij} denote the actual observed number of individuals in category i choosing issue j , and $\hat{Y}_{ij}$ the corresponding estimated counts obtained by each algorithm. The *Error Index* quantifies the percentage of individuals incorrectly

Table 1: R Packages used in the study.

Package	Description	Model features	Computational complexity
lphom[12]	Linear programming under mild homogeneity assumptions. Deterministic estimation of transition proportions.	Prior info via linear constraints or initial global estimates.	Low to moderate.
	Includes functions: <code>tslphom</code> , <code>nsllphom</code> , <code>lclphom</code> , <code>rsllphom</code> .	Helps users evaluate how stable estimated proportions are when data changes.	Increases with local/iterative methods.
	Addresses uncertainty through sensitivity analysis and bootstrap.	–	Efficient for large datasets.
eiCircles[3]	Overdispersed-multinomial models.	Allows unit-level covariates.	Medium to high.
	Assumes voters are grouped into clusters.	Provides estimates of uncertainty.	Depends on number of units/covariates.
eiPack[9]	Hierarchical Bayesian models.	Uses informative and non-informative priors.	High.
	Transition probabilities are estimated via MCMC-derived posterior distributions.	Increases flexibility and enables probabilistic outcome interpretations.	Intensive for large datasets.
	Provides estimates of posterior distributions.	Quantifies uncertainty and provides comprehensive insights into potential scenarios.	–
ecolRxC[14]	Latent structure model focusing on individual preferences and issue positions.	Allows causal interpretation.	High.
	Uses discrete choice models.	Explains political behavior by basing its algorithm on voter preferences to predict choices.	Demanding in high-dimensional settings.
eiopt2[10]	Quadratic optimization minimizing distances between unit and global profiles.	Prior info via numerical constraints.	Moderate to high.
	Solves constrained problems using <code>alabama</code> and <code>quadprog</code> .	–	Depends on dimension and constraints.

assigned to the different categories. This index is formally defined as:

$$\text{Error Index} = 100 \times \frac{0.5 \cdot \sum_{i=1}^R \sum_{j=1}^C |Y_{.ij} - \hat{Y}_{.ij}|}{\sum_{i=1}^R \sum_{j=1}^C Y_{.ij}}.$$

The Error Index ranges from 0, indicating perfect accuracy, to 100, representing the maximum possible discrepancy. Having discussed algorithms and accuracy measures, the next step is applying these approaches to a real-world case study to evaluate their ability to accurately estimate voter transitions.

4 Computation of Transition Matrices and Analysis: 2015 Spanish General Election

4.1 Data and Empirical Framework

This study uses data collected from two pre- and post-electoral surveys conducted by the CIS in connection with the Spanish General Elections of December 20th, 2015. The pre-electoral survey includes responses from 17,452 voters, while the post-electoral survey provides information on only 6,242 of the voters interviewed in the pre-electoral study. These surveys offer a detailed view of voters, including their political and social opinions, voting habits, and electoral history. The contribution analyzes the shift from voting intention (pre-electoral study) to actual voting behavior (post-electoral study). This dataset provides an ideal testing environment for EI algorithms, as it allows direct comparison between predictions and reported outcomes, assessing their accuracy and reliability, given its panel structure. The findings are especially relevant for cases where full information on intention-to-vote transitions is unavailable. Notice when estimating voter transition matrices using paired samples (intended versus actual votes), each EI method relies on specific assumptions and practical scenarios, Table 2 shows them.

Table 2: Hypotheses and assumptions for voter transition matrix estimation.

Package	Assumptions	Practical Use
<code>lphom</code>	Homogeneity of transitions across units.	In general situations and with estimates of prior voter transitions.
<code>eiCircle</code>	Voters grouped into homogeneous groups.	Suitable when dealing with homogeneous districts.
<code>eiPack</code>	Hierarchical modeling of voter transitions.	Useful for uncertainty quantification and when prior knowledge is available.
<code>ecolRxC</code>	Symmetric voter transitions.	Recommendable when an underlying political theory is required.
<code>eiopt</code>	Numerical constraints on voter behavior.	When joint estimation of both local and global transitions is needed.

The analysis excludes incomplete responses and votes for regional parties. However, invalid ballots (IB), blank ballots (BB), abstentions (ABS), “don’t know” responses (DK), and “decline to answer” responses (DA) are included. The final dataset analyzed comprises 5,508 voters.

Table 3 shows the discrepancy between declared voting intentions and actual votes, revealing shifts in support between parties. This highlights the complexities of electoral dynamics, potentially driven by strategic voting, campaign events, or social pressures. Most respondents maintained their initial voting preference, but many vote transfers came from those who initially answered “Don’t Know” in the Pre-electoral survey. Some who intended to vote

for C's ended up choosing PP, PSOE, or Podemos, while others shifted from PSOE and IU to Podemos, showing a realignment of voter preferences.

Table 3: Voting intention and declarations in pre- and post-electoral studies.

		Vote declaration										
		PP	PSOE	IU	UPyD	C's	Podemos	IB	BB	Abs	DA	Total
Voting Intention	PP	846	29	3	0	38	5	0	7	102	58	1088
	PSOE	42	702	13	0	17	68	2	8	78	52	982
	IU	3	19	97	0	5	55	1	1	6	16	203
	UPyD	1	1	0	2	1	2	0	1	1	0	9
	C's	78	72	12	4	393	61	3	6	45	39	713
	Podemos	4	24	18	0	9	471	1	0	37	19	583
	IB	2	2	3	0	6	3	10	2	4	2	34
	BB	23	17	3	2	20	16	4	27	41	16	169
	Abs	43	31	7	1	27	34	2	7	264	39	455
	DK	152	170	34	5	111	145	8	17	138	218	998
	DA	35	39	6	0	15	28	1	3	28	119	274
	Total	1229	1106	196	14	642	888	32	79	744	578	5508

Figure 1 provides the transition matrix derived from these data, i.e., the real proportions of voters who reported a voting intention for a particular row and declared a vote for the option in the column. The transition matrix reveals voter movement between the intended vote and the actual vote. High loyalty exists within PP (77.8%) and PSOE (71.5%). IU voters significantly shift to Podemos (27.1%). C's voters show loyalty (55.1%) but also move to PP (10.9%). Most intenders of abstention remain so (58.0%), while others shift to PP or PSOE. Intenders to cast an Invalid or a Blank ballot often shift to larger parties.

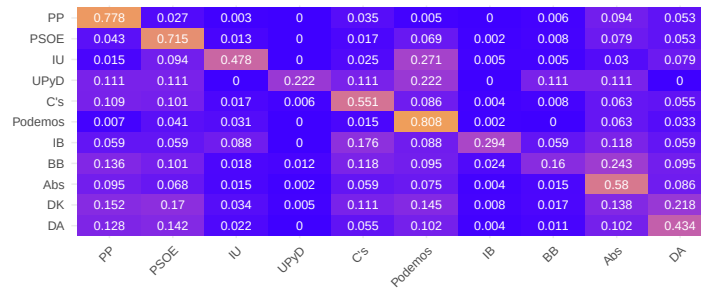


Fig. 1: Transition matrix with actual responses.

4.2 Main Findings

Figure 2 displays transition matrices for methods without prior voter behavior information— i.e. considering only the information in the Total row and Total column of Table 3—. `lphom`, `eiOpt2`, and `eiCircle` show similar estimations, generally underestimating shifts except for UPyD and undecided voters. `eiPack` overestimates PP voter loyalty, while `ecoRx` underestimates loyalty across all voter responses.

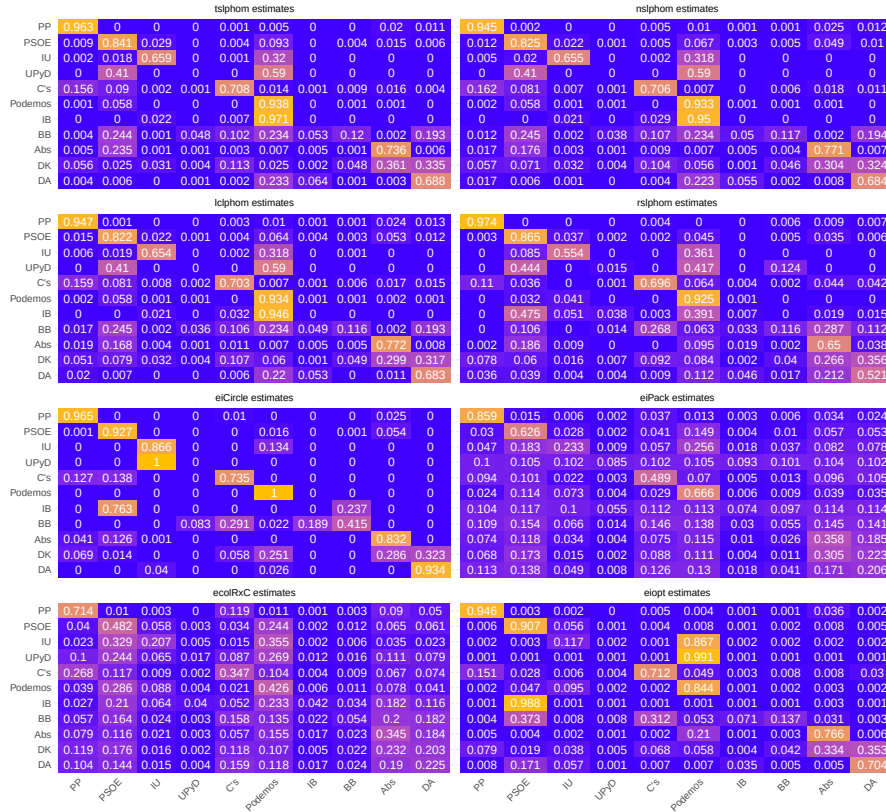


Fig. 2: Transition matrices without prior information.

When incorporating prior information— that is, the transition matrix reported in Figure 1—, `lphom` and `eiPack` overestimate voter loyalty to PP, while `tsIphom` and `rsIphom` functions provide estimates close to actual values. The other methods estimates vary, but are closer to actual values compared to when prior information is absent. This observation is exemplified by the graphical depiction of the Error Index, as presented in Figure 4. The figure distinctly differentiates between methodologies employing prior information (denoted in red) and those

that do not (denoted in blue). Methodologies lacking prior information show elevated error rates relative to those incorporating such information. Among the methodologies devoid of prior information **eiPack** provides the optimal fit, whereas the **tslphom** algorithm attains the minimal Error Index value when prior information is incorporated.

tslphom estimates											rslphom estimates										
PP	0.806	0.015	0.001	0	0.04	0.011	0.001	0.007	0.076	0.044	0.885	0.001	0	0	0.044	0.015	0	0.001	0.028	0.024	
PSOE	0.044	0.731	0.012	0	0.013	0.074	0.002	0.006	0.072	0.046	0.042	0.766	0.014	0.002	0.002	0.101	0.007	0.008	0.047	0.012	
IU	0.013	0.099	0.479	0	0.021	0.27	0.005	0.005	0.03	0.079	0.004	0.119	0.492	0.001	0.006	0.269	0.003	0.005	0.03	0.072	
UPyD	0.111	0.111	0	0.222	0.111	0.222	0	0.111	0.111	0	0.111	0.111	0	0.222	0.111	0.222	0	0.111	0.111	0	
C's	0.124	0.098	0.016	0.005	0.555	0.078	0.004	0.008	0.063	0.055	0.114	0.112	0.019	0.005	0.954	0.062	0.008	0.003	0.069	0.054	
Podemos	0.004	0.043	0.036	0.002	0.009	0.815	0.001	0	0.06	0.03	0.001	0.054	0.051	0.002	0	0.838	0.002	0	0.026	0.006	
IB	0.067	0.064	0.088	0	0.175	0.088	0.283	0.058	0.118	0.059	0.125	0.087	0.094	0	0.164	0.082	0.202	0.056	0.131	0.059	
BB	0.138	0.101	0.017	0.011	0.118	0.095	0.023	0.159	0.243	0.095	0.158	0.104	0.015	0.005	0.115	0.095	0.02	0.15	0.243	0.095	
Abs	0.094	0.066	0.015	0.002	0.057	0.072	0.005	0.016	0.581	0.092	0.086	0.052	0.013	0.001	0.039	0.059	0.009	0.02	0.613	0.109	
DK	0.111	0.168	0.036	0.005	0.116	0.137	0.008	0.019	0.166	0.235	0.03	0.134	0.025	0.004	0.137	0.101	0.003	0.025	0.244	0.288	
DA	0.13	0.142	0.02	0	0.054	0.101	0.004	0.012	0.102	0.434	0.151	0.14	0.009	0	0.053	0.092	0.003	0.014	0.103	0.434	
lsiphom estimates											rslphom estimates										
PP	0.903	0.002	0	0	0.041	0.012	0	0.002	0.016	0.022	0.817	0.012	0.001	0	0.035	0.002	0	0.006	0.08	0.048	
PSOE	0.041	0.767	0.006	0.002	0.002	0.102	0.007	0.009	0.062	0.001	0.031	0.734	0.016	0	0.014	0.071	0.002	0.01	0.071	0.05	
IU	0.002	0.129	0.504	0	0.002	0.264	0.002	0.005	0.032	0.06	0.004	0.094	0.489	0	0.013	0.328	0.004	0.002	0.018	0.048	
UPyD	0.111	0.111	0	0.222	0.111	0.222	0	0.111	0.111	0	0.04	0.139	0	0.173	0.154	0.401	0	0.034	0.052	0.006	
C's	0.096	0.122	0.019	0.005	0.559	0.05	0.01	0.001	0.072	0.065	0.136	0.086	0.011	0.004	0.568	0.06	0.004	0.009	0.066	0.054	
Podemos	0.001	0.053	0.05	0.002	0	0.886	0.003	0	0.003	0.002	0.002	0.046	0.044	0	0.009	0.828	0.002	0	0.046	0.023	
IB	0.167	0.098	0.088	0	0.156	0.077	0.146	0.052	0.158	0.059	0.025	0.174	0.083	0.002	0.169	0.13	0.231	0.037	0.086	0.065	
BB	0.176	0.106	0.013	0.003	0.112	0.096	0.018	0.139	0.243	0.096	0.081	0.091	0.014	0.013	0.14	0.092	0.024	0.16	0.264	0.121	
Abs	0.064	0.049	0.008	0	0.027	0.056	0.008	0.02	0.637	0.13	0.051	0.078	0.014	0.002	0.052	0.085	0.009	0.016	0.593	0.101	
DK	0.025	0.123	0.034	0.005	0.145	0.1	0.003	0.027	0.241	0.298	0.144	0.171	0.03	0.006	0.11	0.127	0.008	0.017	0.163	0.224	
DA	0.165	0.141	0.006	0.001	0.051	0.084	0.003	0.012	0.106	0.431	0.104	0.131	0.02	0	0.051	0.126	0.007	0.012	0.104	0.445	
eiPack estimates																					
PP	0.806	0.015	0.001	0	0.04	0.011	0.001	0.007	0.076	0.044	0.885	0.001	0	0	0.044	0.015	0	0.001	0.028	0.024	
PSOE	0.044	0.731	0.012	0	0.013	0.074	0.002	0.006	0.072	0.046	0.042	0.766	0.014	0.002	0.002	0.101	0.007	0.008	0.047	0.012	
IU	0.013	0.099	0.479	0	0.021	0.27	0.005	0.005	0.03	0.079	0.004	0.119	0.492	0.001	0.006	0.269	0.003	0.005	0.03	0.072	
UPyD	0.111	0.111	0	0.222	0.111	0.222	0	0.111	0.111	0	0.111	0.111	0	0.222	0.111	0.222	0	0.111	0.111	0	
C's	0.124	0.098	0.016	0.005	0.555	0.078	0.004	0.008	0.063	0.055	0.114	0.112	0.019	0.005	0.954	0.062	0.008	0.003	0.069	0.054	
Podemos	0.004	0.043	0.036	0.002	0.009	0.815	0.001	0	0.06	0.03	0.001	0.054	0.051	0.002	0	0.838	0.002	0	0.026	0.006	
IB	0.067	0.064	0.088	0	0.175	0.088	0.283	0.058	0.118	0.059	0.125	0.087	0.094	0	0.164	0.082	0.202	0.056	0.131	0.059	
BB	0.138	0.101	0.017	0.011	0.118	0.095	0.023	0.159	0.243	0.095	0.158	0.104	0.015	0.005	0.115	0.095	0.02	0.15	0.243	0.095	
Abs	0.094	0.066	0.015	0.002	0.057	0.072	0.005	0.016	0.581	0.092	0.086	0.052	0.013	0.001	0.039	0.059	0.009	0.02	0.613	0.109	
DK	0.111	0.168	0.036	0.005	0.116	0.137	0.008	0.019	0.166	0.235	0.03	0.134	0.025	0.004	0.137	0.101	0.003	0.025	0.244	0.288	
DA	0.13	0.142	0.02	0	0.054	0.101	0.004	0.012	0.102	0.434	0.151	0.14	0.009	0	0.053	0.092	0.003	0.014	0.103	0.434	

Fig. 3: Transition matrices with prior information.

5 Conclusions

The findings highlight that no single approach provides a definitive solution to the challenges of ecological inference. Some approaches prioritize interpretability and theoretical robustness, while others emphasize computational speed or optimization techniques to refine estimates. Uncertainty remains a central issue, as each method incorporates it differently, affecting both the reliability and practical applicability of the results. These differences underscore the importance of handpicking a technique based on the specific electoral context and data characteristics. The empirical analysis based on real-world electoral data confirms these observations, demonstrating how methodological choices impact the interpretation of voter transitions. Overall, the study shows computational models can enhance our understanding of voter behavior but must be used with

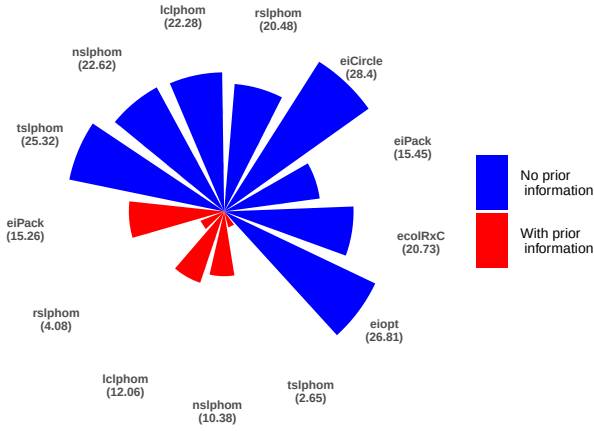


Fig. 4: Error Index across the different used algorithms.

caution, acknowledging the inherent limitations of inference from aggregate data. This study contributes to the development of decision support methods for analyzing electoral behavior under uncertainty. Through ecological inference, it provides tools to recover individual-level patterns from aggregated data, improving the interpretation of collective decisions in complex scenarios. Future advancements should aim to refine these methodologies, incorporating more flexible uncertainty modeling and improving computational performance to better align with the complexity of real-world electoral decisions

Acknowledgments. Pérez-Asurmendi acknowledges financial support from the Fundación Areces (Grant CISP18A6216). Pérez-Asurmendi and de Andrés Calle acknowledge funding from project PID2022-139469NB-I00, funded by Spanish MCIN/AEI/10.13039/501100011033 and ERDF. Pavía acknowledges funding from Generalitat Valenciana (project CIAICO/2023/031) and MCIN/AEI/10.13039/501100011033 (project PID2021-128228NB-I00).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bavard, S., Stuchlý, E., Konovalov, A., Gluth, S.: Humans can infer social preferences from decision speed alone. *PLOS Biology* **22**, e3002686 (2024)
2. Brown, P.J., Payne, C.D.: Aggregate data, ecological regression, and voting transitions. *Journal of the American Statistical Association* **81**, 452–460 (1986)
3. Forcina, A., Pavía, J.M.: eiCircles: Ecological inference of RxC tables by overdispersed-multinomial models (2023), <https://cran.r-project.org/web/packages/eiCircles>

4. Forcina, A., Pavía, J.M.: From Corrado Gini's early contributions to overdispersion to modern models of voting behaviour. *Statistical Science* **0**(00), 1–10 (2024). <https://doi.org/10.1214/24-ST960>, forthcoming
5. Goodman, L.A.: Ecological regressions and behavior of individuals. *American Sociological Review* **18**(6), 663–664 (1953)
6. Greiner, D.J., Quinn, K.M.: $R \times C$ ecological inference: Bounds, correlations, flexibility and transparency of assumptions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(1), 67–81 (2009)
7. King, G.: *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press (1997)
8. Klima, A., Thurner, P.W., Molnar, C., Schlesinger, T., Küchenhoff, H.: Estimation of voter transitions based on ecological inference: An empirical assessment of different approaches. *AStA-Advances in Statistical Analysis* **100**, 133–159 (2016)
9. Lau, O., Moore, R., Kellermann, M.: eiPack: Ecological inference and higher-dimension data management (2023), <https://cran.r-project.org/package=eiPack>
10. Pavía, J.M., Fernández, V.: eiOpt2: Ecological inference for $R \times C$ tables via nonlinear quadratic optimization (2025), <https://cran.r-project.org/package=eiOpt2>
11. Pavía, J.M., Romero, R.: Improving estimates accuracy of voter transitions. Two new algorithms for ecological inference based on linear programming. *Sociological Methods & Research* **53**(3), 1491–1533 (2024)
12. Pavía, J.M., Romero, R.: lphom: Ecological inference by linear programming under homogeneity (2024), <https://cran.r-project.org/web/packages/lphom>
13. Pavía, J.M., Romero, R.: Symmetry estimating $R \times C$ vote transfer matrices from aggregate data. *Journal of the Royal Statistical Society Series A: Statistics in Society* **187**(4), 919–943 (2024)
14. Pavía, J.M., Thomsen, S.R.: ecolRxC: Ecological inference of $R \times C$ tables by latent structure approaches (2023), <https://cran.r-project.org/web/packages/ecolRxC>
15. Pavía, J.M., Thomsen, S.R.: ecolRxC: Ecological inference estimation of $R \times C$ tables using latent structure approaches. *Political Science Research and Methods* **0**(0), 1–19 (2024). <https://doi.org/10.1017/psrm.2024.57>, online available
16. Plescia, C., De Sio, L.: An evaluation of the performance and suitability of $R \times C$ methods for ecological inference with known true values. *Quality & Quantity* **52**, 669–683 (2018)
17. Robinson, W.: Ecological correlations and the behavior of individuals. *American Sociological Review* **15**(3), 351–357 (1950)
18. Romero, R., Pavía, J.M., Martín, J., Romero, G.: Assessing uncertainty of voter transitions estimated from aggregated data. Application to the 2017 French presidential election. *Journal of Applied Statistics* **47**(13-15), 2711–2736 (2020)
19. Rosen, O., Jiang, W., King, G., Tanner, M.A.: Bayesian and frequentist inference for ecological inference: The $R \times C$ case. *Statistica Neerlandica* **55**(2), 134–156 (2001)
20. Selb, P., Munzert, S.: Voter overrepresentation, vote misreporting, and turnout bias in postelection surveys. *Electoral Studies* **32**(1), 186–196 (2013)
21. Thomsen, S.R.: *Danish elections, 1920–79: A Logit Approach to Ecological Analysis and Inference*. Politica, Aarhus (1987)