SORT 47 (1) January-June 2023, 151-186 DOI: 10.57645/20.8080.02.4

Data wrangling, computational burden, automation, robustness and accuracy in ecological inference forecasting of R×C tables

Jose M. Pavía[1] and Rafael Romero[2]

Abstract

This paper assesses the two current major alternatives for ecological inference, based on a multinomial-Dirichlet Bayesian model and on mathematical programming. Their performance is evaluated in a database made up of almost 2000 real datasets for which the actual cross-distributions are known. The analysis reveals both approaches as complementarity, each one of them performing better in a different area of the simplex space, although with Bayesian solutions deteriorating when the amount of information is scarce. After offering some guidelines regarding the appropriate contexts for employing each one of the algorithms, we conclude with some ideas for exploiting their complementarities.

MSC: *90C05, 62F15, 62Q05, 05A17, 91-08, 05-04.*

Keywords: *Ecological inference; Voter transitions; US voting rights; two-way contingency tables; ei.MD.bayes; lphom; R-packages.*

1. Introduction

Ecological inference forecasting aims to estimate the inner-cells values of a set of related R×C contingency tables when only the margins are known. Ecological inference is a particular instance of cross-level inference. In ecological inference, the objective is to infer individual-level behavior from aggregate-level (i.e., ecological) data when individual-level data are not available. This outlines one of the more conspicuous and

[1] GIPEyOP, Universitat de Valencia, Valencia (Spain), email: pavia@uv.es ORCID: https://orcid.org/0000-0002-0129-726X

[2] Universidad Politécnica de Valencia, Valencia (Spain), email: rromero@eio.upv.es

Received: September 2022
Accepted: April 2023

long-standing problems of social sciences present in many disciplines, from marketing and epidemiology to sociology and political science, and encompassing geography, economics and quantitative history (King, 1997; Petropoulos et al., 2022). In ecological inference, the problem arises because information is lost when aggregating across individuals, the fundamental challenge being that many different possible relationships at the individual level can produce the same observations at the aggregate level.

Despite the dangers of cross-level inferences being widely acknowledged, arising from the so-called group or ecological fallacy (e.g., Allport, 1924; Robinson, 1950) and the Simpson paradox (e.g., Gehlke and Biehl, 1934; Simpson, 1951), the solutions promised by this approach soon attracted the interest of researchers, mainly within the discipline of political science (Ogburn and Goltra, 1919; Gosnell and Gill, 1935). A particularly relevant instance of this problem arises when the focus is on estimating/forecasting the inner-cells values of a set of related R×C contingency tables when only the margins are known. For example, finding out from the data available on a set of voting units (e.g., counties or precincts) how different people (grouped, for instance, according to their religion: Catholics, Protestants, Muslims, agnostics, ...) split their votes among different candidates, or estimating the vote transfers between two elections. Focusing on the second example, the objective is to ascertain the cross-tabulated distribution of votes in each unit and in the whole electoral space by just using the sets of votes recorded in the units in the two elections (the margins of the tables).

The fundamental challenge of the ecological inference forecasting problem lies in the fact that there are a multitude of ways to determine the interior cell counts of a table with the same aggregated margins, and this indeterminacy cannot be solved collecting data from more units (Manski, 2007; Greiner and Quinn, 2009; Forcina and Pellegrino, 2019). To disentangle this, a basic assumption of similarity (and, sometimes, the use of covariates) is routinely considered. The aim of this paper is to assess in terms of accuracy, robustness and simplicity, and also considering the features of computational burden, automation and data wrangling requirements, the two main alternatives for ecological inference forecasting available in the R packages `eiPack` (Lau, Moore and Kellermann, 2020) and `lphom` (Pavía and Romero, 2021).

Klima et al. (2016) and Plescia and De Sio (2018), working independently and after analyzing the main methods developed up to that moment, conclude that the algorithm programmed in the `ei.MD.bayes` function of the `eiPack` package is the one that generates the best solutions. However, Romero et al. (2020) and Pavía and Romero (2022) have recently proposed three new algorithms (`lphom`, `tslphom` and `nslphom`), available in the `lphom` package (Pavía and Romero, 2021), whose performance seems to exceed, at least in certain circumstances, the estimates achieved with `ei.MD.bayes`. Romero et al. (2020) and Romero and Pavía (2021) report, when studying the vote transfers between the first and second rounds of the 2017 French presidential elections, that `ei.MD.bayes` produces unusable solutions when working with limited voting units.

Specifically, Romero and Pavía (2021) find when working with outcomes at the regional level (13 voting units) and with outcomes at the department level (108 units) that ei.MD.bayes generates solutions without socio-political sense. They obtain this result both when using the default options of ei.MD.bayes and when tuning the parameters of the function. Only when working with outcomes at the district level (577 units), and after tuning the model parameters (incurring significant data analysis and computational costs), are they able to achieve satisfactory solutions. These findings contrast, on the one hand, with the excellent solutions that ei.MD.bayes provides, using its default options, for the dataset (212 units) included in the eiPack package and with the conclusions reported in Klima et al. (2016) and Plescia and De Sio (2018) and, on the other hand, with the satisfactory solutions that are achieved, in a few seconds and without specification costs, using the default options of the lphom functions. Therefore, a broad and systematic study of comparison is needed between the functions of both packages to determine the empirical strengths and weaknesses of each algorithm and the circumstances in which each of them will generate better estimates.

Although a significant part of the studies of this nature use simulated datasets to assess the quality of the estimates (e.g., Ferree, 2004; Greiner and Quinn, 2010; Klima et al., 2016; Klein, 2019; Klima et al., 2019; Martín, 2020; Barreto et al, 2022) since the data of interest in real situations is usually unknown (indeed, this is the purpose of the different procedures), in this research we use real data for the assessments. This is in line with Plescia and De Sio (2018) and Pavía and Romero (2022) and is the approach recommended by Collingwood et al. (2016, p. 93), who "argue that real election data should be considered in a side-by-side comparison". In particular, the performance of the different algorithms is evaluated by exploiting the data from a singular database made up of almost 500 elections for which the current cross-distribution of votes in the entire electoral space is known. This database includes all the elections analyzed in Plescia and De Sio (2018) and Pavía and Romero (2022).

The assessment of the algorithms will not only focus on evaluating their accuracy in predicting the cross distributions but also on other considerations such as their data wrangling and specification requirements. On the one hand, the procedure implemented in ei.MD.bayes is a complex procedure, based on Markov chain Monte Carlo (MCMC) methods, that (i) demands the specification and tuning of a large number of parameters (among them, a priori distributions with their hyperparameters, the number of initial iterations to be discarded, the length of the chains or the jump between accepted values in each chain) and (ii) requires, before using the function, an intensive data pre-processing to guarantee the congruence between the marginal distributions of rows and columns of each table. On the other hand, the procedures implemented in lphom, based on mathematical programming, can negotiate different scenarios in terms of (lack of) congruence between marginal distributions and only require, in the nslphom algorithm, specification of the number of iterations. All these issues must be weighed up when choosing an algorithm to solve a problem.

Given that in real situations the inner-cells of the contingency tables are generally unknown – at most, we can check the solutions for their plausibility but not the quality (accuracy) of the predictions – we also evaluate the robustness and sensitivity of the different algorithms in either more stressful or simpler scenarios. Starting from the observed database composed of 493 elections, we construct new sets of electoral results by aggregating voting units and/or voting options. This will allow the scenarios under analysis to be increased and the algorithms to be evaluated in new situations, where the problem is simplified (with fewer cells in the transfer matrices) and/or with less data available (with fewer voting unit observations). In total, using real data at all times, we analyze the equivalent of 1972 elections.

The rest of the document is structured as follows. The second section details the characteristics of the methods implemented in both functions. The third section is dedicated to data. The fourth section compares and analyses the results attained after applying `ei.MD.bayes`, with different specifications, and the `lphom` algorithms to the initial datasets corresponding to the 493 elections. The fifth section explores the robustness and sensitivity of the estimates in the new scenarios, created from the base data. Section 6 reviews the analysis and, by pooling the results of all the datasets, looks at, among other issues, the features that affect the accuracy of the estimates in both approaches. Finally, Section 7 summarizes the findings, states some recommendations and suggests directions for further research. This paper is complement with two files with Supplementary Material.

2. The methods

In the ecological inference forecasting problem, the units of analysis are contingency tables with observed row and column marginals and the objective is to estimate the unobserved internal cells for each unit (and/or for the aggregation of all the tables). Mathematically, denoting by $i = 1, 2, ..., I$ the index for the units, $j = 1, 2, ..., R$ the index for the rows and $k = 1, 2, ..., C$ the index for the columns (where I, R and C represent, respectively, the number of units, rows and columns), the problem can be stated, as expressed in Table 1, as one of estimating the (red) values N_{jki} $\forall i, j, k$, given their row and column aggregations: $N_{j.i} = \sum_k N_{jki}$ and $N_{.ki} = \sum_j N_{jki}$ (where $N_{..i} = \sum_{jk} N_{jki} = \sum_j N_{j.i} = \sum_k N_{.ki}$).

Table 1. *A typical $R \times C$ unit in ecological inference. Red quantities are the unobserved counts.*

	col_1	...	col_k	...	col_C	
row_1	N_{11i}	...	N_{1ki}	...	N_{1Ci}	$N_{1.i}$
...	...	...	...	...	...	...
row_j	N_{j1i}	...	N_{jki}	...	N_{jCi}	$N_{j.i}$
...	...	...	...	...	...	...
row_R	N_{R1i}	...	N_{Rki}	...	N_{RCi}	$N_{R.i}$
	$N_{.1i}$	...	$N_{.ki}$	...	$N_{.Ci}$	$N_{..i}$

Many algorithms for solving the ecological inference forecasting problem can be found in the literature. In this research, the estimates obtained from two procedures with different philosophical and mathematical substrates are compared: on the one hand, the three algorithms implemented in the `lphom` package (Pavía and Romero, 2021) and, on the other hand, several specifications of the procedure available in the `ei.MD.bayes` function of the `eiPack` package (Lau et al., 2020). The first algorithms are based on mathematical programming, while the second procedure has its roots in Bayesian statistics. Other methods to solve this problem include the iterative version of the 2×2 model proposed by King (see King, 1997; Imai, King and Lau, 2008; Collingwood et al., 2016; Choirat et al., 2017), the aggregated compound multinomial model proposed by Brown and Payne (1986) or the generalization of the Goodman regression method (see Goodman, 1953, 1959; Collingwood et al., 2016).

Despite the different foundations of the various procedures, they all rely on the same information sources and basic assumptions to obtain their estimates. All of them exclusively use the information contained in the margins of the tables and assume a hypothesis of similar behavior between different units to overcome the problems of identifiability and indeterminacy. In particular, `lphom` assumes small distances across units among p^i_{jk} and also with p_{jk} and `ei.MD.bayes` considers that, conditional on the row, j, all the p^i_{jk} of the different units are realizations of a common probability distribution, where $p^i_{jk} = N_{jki}/N_{j.i}$ and $p_{jk} = \sum_i N_{jki}/\sum_i N_{j.i}$ are, respectively, the (unknown) unit and global cell fractions. Both procedures also impose (explicitly `lphom` and implicitly `ei.MD.bayes`) the restrictions that are derived from the available information. The unit cell fractions, p^i_{jk}, that both approximations estimate must be compatible with the marginals of each unit and of the set of tables.

2.1. The model in ei.MD.bayes

The procedure implemented in the `ei.MD.bayes` function uses a method based on a hierarchical Multinomial-Dirichlet model initially proposed for 2×2 tables by King, Rosen and Tanner (1999) and later generalized for R×C tables by Rosen et al. (2001). Specifically, denoting the row marginal and the column marginal fractions of unit i by, respectively, $X_{ji} = N_{j.i}/N_{..i}$ and $T_{ki} = N_{.ki}/N_{..i}$, the hierarchical Multinomial-Dirichlet model, without covariates, proposed by Rosen et al. (2001) assumes, for the first level of the hierarchy, that the vector of column marginal counts in unit i follows a Multinomial distribution of the form:

$$(N_{.1i},...,N_{.ki},...,N_{.Ci}) \sim \text{Multinomial}(N_{..i}, \sum_{j=1}^{R} p^i_{j1}X_{ji},..., \sum_{j=1}^{R} p^i_{jk}X_{ji},..., \sum_{j=1}^{R} p^i_{jC}X_{ji})$$

and, for the second level of the hierarchy, that the vector of cell fractions for row j ($j = 1,\ldots,R$) in unit i ($i = 1,\ldots,I$) follows a Dirichlet distribution with C parameters, constant across units:

$$(p^i_{j1},\ldots,p^i_{jk},\ldots,p^i_{jC}) \sim \text{Dirichlet}(\alpha_{j1},\ldots,\alpha_{jk},\ldots,\alpha_{jC})$$

where the prior on each α_{jk} is assumed to be:

$$\alpha_{jk} \sim \text{Gamma}(\lambda_1, \lambda_2)$$

The first level of the hierarchy introduces the information of the margins by modelling, conditional on the observed row totals, the observed column totals as multinomial distributions independent across units. The second level of the hierarchy enables the borrowing of strength across the estimates of the (unobserved) row-cell proportions/fractions of different units by modelling them as Dirichlet distributions independent across rows and conditional independent across units. The third level of the hierarchy considers a fairly non-informative distribution for the Dirichlet parameters. The hierarchical model not only increases efficiency (decreases variation) of the estimates by borrowing statistical strength across units, but it also makes it possible to obtain estimates of the unobserved quantities p^i_{jk}.

This hierarchical Bayesian model is fit by `ei.MD.bayes` using a Metropolis-within-Gibbs algorithm (Robert and Casella, 2004). Conducting an analysis employing this model involves two steps: first, calibrating priors and tuning parameters used for Metropolis-Hastings sampling and, second, generating proper MCMC draws. This requires analysts highly trained in Bayesian statistics since, in addition to the need to tune a large number of parameters, assessing convergence of MCMC chains tends to be difficult is this setting (Rosen et al., 2001; Lau, Moore, and Kellermann, 2007): sometimes the scarce information available in the margins of the tables (i.e., regarding p^i_{jk} bounds) can lead to extremely slow mixing of MCMC chains. Furthermore, when the number of units is scarce and all the margins of the unit tables are sufficiently populated, some substantive knowledge of the phenomenon under study is also required to properly customize prior hyperparameters. As Wakefield (2004) notes, the inherent problems of identifiability and indeterminacy that characterizes ecological inference is likely to lead to solutions sensitive to the choice of prior so, as Lau et al. (2007, p. 46) recommend, "[u]sers should experiment with different assumptions about the prior distribution of the upper-level parameters in order to gauge the robustness of their inference". It is also necessary to properly set issues such as the length of the burn-in period, the thinning parameter and the total length of the chains. It is essential to generate enough iterations for the Markov Chain to converge, as only if a convergence occurs can the samples from a Markov Chain be used in a Monte Carlo integration.

2.2. The model in lphom

The methods included in `lphom`, acronym for "**L**inear **P**rogram model based on the **HOM**ogeneity hypothesis", estimate the p^i_{jk} by solving two sequential linear programs that, conforming to the observed marginal counts, minimizes the L_1 distance of the cell fractions across units. The `nslphom` algorithm (Pavía and Romero, 2022) is an iterative procedure that yields the `lphom` and the `tslphom` solutions as by-products. In its simplest specification, `nslphom` uses equations (1) to (15) to attain its solution. In its step zero, the algorithm solves the basic `lphom` system (Romero et al., 2020) defined by equations (1) to (5).

$$p_{jk} \geq 0 \quad \text{for } j = 1,\ldots,R,\ k = 1,\ldots,C \tag{1}$$

$$\sum_{k=1}^{C} p_{jk} = 1 \quad \text{for } j = 1,\ldots,R \tag{2}$$

$$\sum_{j=1}^{R} \left(\sum_{i=1}^{I} N_{j\cdot i} \right) p_{jk} = \sum_{i=1}^{I} N_{\cdot ki} \quad \text{for } k = 1,\ldots,C \tag{3}$$

$$e_{ik} = N_{\cdot ki} - \sum_{j=1}^{R} N_{j\cdot i} p_{jk} \quad \text{for } k = 1,\ldots,C,\ i = 1,\ldots,I \tag{4}$$

$$\textit{minimize} \sum_{i,k} |e_{ik}| \tag{5}$$

This step zero produces an initial solution matrix $\hat{\mathbf{P}}_0$ = $\left[{}_0\hat{p}_{jk}\right]$ of the matrix, $\mathbf{P} = \left[p_{jk}\right]$, of global cell fractions that is used to start the iterative process that characterizes `nslphom`. In the next steps, for $l = 1,...,ns$ (where ns is the number of steps), the algorithm generates estimates of the unit cell fractions, p^i_{jk}, and the global cell fractions, p_{jk}, by recursively updating the ${}_l\hat{p}_{jk}$ estimates and solving the two sequential systems defined by expressions (6) to (13).

$$p^i_{jk} \geq 0 \quad \text{for } j = 1,\ldots,R,\ k = 1,\ldots,C,\ i = 1,\ldots,I \tag{6}$$

$$\sum_{k=1}^{C} p^i_{jk} = 1 \quad \text{for } j = 1,\ldots,R,\ i = 1,\ldots,I \tag{7}$$

$$\sum_{j=1}^{R} N_{j\cdot i} p^i_{jk} = N_{\cdot ki} \quad \text{for } k = 1,\ldots,C,\ i = 1,\ldots,I \tag{8}$$

$$\varepsilon^i_{jk} = ({}_{l-1}\hat{p}_{jk} - p^i_{jk}) N_{j\cdot i} \quad \text{for } j = 1,\ldots,R,\ k = 1,\ldots,C,\ i = 1,\ldots,I \tag{9}$$

$$\textit{minimize } Z = \sum_{j,k} |\varepsilon^i_{jk}| \quad \text{for } i = 1,\ldots,I \tag{10}$$

$$Z = \sum_{j,k} |\varepsilon^i_{jk}| \quad \text{for } i = 1,\ldots,I \tag{11}$$

$$p^i_{jk} = ({}_{l-1}\hat{p}_{jk} + \delta^i_{jk}) \quad \text{for } j = 1,\ldots,R,\ k = 1,\ldots,C,\ i = 1,\ldots,I \tag{12}$$

$$\textit{minimize} \sum_{j,k} |\delta^i_{jk}| \quad \text{for } i = 1,\ldots,I \tag{13}$$

where ${}_l\hat{p}_{jk}$ is computed by equation (14) using the l-step solutions $\hat{p}^i_{jk}(l)$ attained after solving equations (6)-(13).

$${}_l\hat{p}_{jk} = \sum_{i=1}^{I} \hat{p}^i_{jk}(l) N_{j.i} / \sum_{i=1}^{I} N_{j\cdot i} \quad \text{for } j = 1,\ldots,R,\ k = 1,\ldots,C \tag{14}$$

During the iterative process, the statistic defined by equation (15), which measures the aggregate distance to homogeneity of the recursive solutions, is also computed. This statistic is utilized to determine the `nslphom` solution, which corresponds to the iteration l^* minimizing (15) .

$$HET_l = 100 \cdot \frac{0.5 \sum_{ijk} \ \hat{p}^i_{jk}(l) N_{j \cdot i} - \ {}_l\hat{p}_{jk} N_{j \cdot i}}{\sum_{ij} N_{j \cdot i}} \tag{15}$$

Once the iterative process has finished, we have three solutions: the `lphom` solution, which corresponds to the step zero solution, the `tslphom` solution, which corresponds to the solution attained in step one and, finally, the solution corresponding to step l^*, which is the `nslphom` solution. Note that the `lphom` solution only provides estimates for the inner-cells of the global table. The above algorithm is quite automatic with only one parameter to tune: the number of steps, *ns*. According to Pavía and Romero (2022), the minimum of equation (15) is usually reached after very few steps. Indeed, the default option of the `nslphom` function considers only ten steps.

3. The data

Given the secret nature of voting, internal cell counts of global and unit tables are mostly unobserved. Sometimes, however, they are available, as when voters cast ballots with several votes in the same ballot and they are counted and published jointly. This is (partially) the case of the New Zealand general elections since 2002 and of the 2007 Scottish Parliamentary election, where a mixed-member election system is employed. In these elections, voters cast two independent votes – one for a list (usually a party list) and another for a local candidate – and the electoral authorities publish/published party-candidate cross-tabulations at district level and marginal results at polling station level. This provides a unique opportunity to assess algorithms by comparing actual observed global cross-tables with forecasted ecological tables. In each district, the `ei.MD.bayes` and `nslphom` functions can be run to forecast the internal cell counts (or fractions) of the district table using as inputs the marginal results at polling station level, to afterwards compare forecasts and actual observed values.

Specifically, we collected 493 datasets composed of marginal polling stations' results and party-candidate cross tables corresponding to the same number of elections (districts): 420 datasets came from the 2002, 2005, 2008, 2011, 2014 and 2017 New Zealand general elections and 73 datasets from the 2007 Scottish Parliament election. In the case of New Zealand, the raw files of the cross-distributions of votes at district level (with parties by rows and candidates by columns) and of the marginal distributions of votes at polling station level were downloaded from the official web page of the electoral commission of New Zealand (www.electionresults.org.nz). In the case of Scotland, the authors gained access to the data via personal communication with Carolina Plescia, who had downloaded the raw files from the Scotland Electoral Office website in 2011. The Scottish data are no longer available on that site.

Before using the data, every election-district dataset is checked for internal consistency and pre-processed in order to guarantee that the accounting equalities $\sum_j N_{j\cdot i} = \sum_k N_{\cdot ki}$ (for $i = 1,\dots,I$) and $\sum_i N_{j\cdot i} = \sum_k N_{jk\cdot}$ and $\sum_i N_{\cdot ki} = \sum_j N_{jk\cdot}$ (for $j = 1,\dots,R$ and $k = 1,\dots,C$) hold in each dataset for, respectively, each polling station (voting unit) and the whole district, where $N_{jk\cdot}$ $(= \sum_i N_{jki})$ are the internal cell counts (observed in these datasets) of the district tables.

In the case of the New Zealand datasets, we have removed: (i) the rows with all their values being zero or non-available in the parties' and candidates' files; and (ii) the row corresponding to the polling unit identified as "Votes allowed for party only" in the parties' files and, equally, the corresponding column ("Party vote only") in the cross-distribution files. The second group of deletions was performed because the voting unit "Votes allowed for party only" has no equivalent in the candidates' files. In addition to these general pre-processing tasks, we merged the voting units identified as "Voting places where less than 6 votes were taken" (row 100) and "Ordinary votes before polling day" (row 101) in the party and candidate files of the 43rd district (Rangitikei) of the 2014 election. We did this to solve a mismatch between both files as the values in their 100th and 101st rows were, respectively, 3 and 2 and 8465 and 8466.

Finally, before starting any analysis and as is common practice when forecasting real tables (e.g., Klima et al., 2016; Plescia and De Sio, 2018; Klein, 2019; Pavía and Aybar, 2020; Pavía and Romero, 2022), we merged very small electoral options. In each dataset, those parties or candidates which individually did not reach at least 3% of the election-district vote were grouped in the option 'Others'. Hereinafter, we call this set of datasets the reference database. Table 2 offers some summary statistics of this database, with more details available in Pavía (2022).

As can be seen in Table 2, we have some variety in terms of the features in the datasets collected. In particular, looking at the last two columns of Table 2, we see that our database also presents an interesting diversity in terms of voters' distribution among cells within rows. And this despite our cross-tables coming from ticket-splitting in concurrent elections, where more cell fractions close to one (zero) are routinely recorded than in other contexts, such as in demographic voting. This, undoubtedly, enriches the analyses by allowing the algorithms to be evaluated in different contexts. Indeed, according to Park, Hanmer and Biggers (2014), gauging the accuracy of ecological inference procedures across different contexts adds robustness to the conclusions, particularly for studying what happens when the across-unit variance varies and/or when the number of units is small.

According to Wakefield (2004), smaller areas are preferable (i.e., voting units with a small number of voters) because it reduces the possibility of ecological bias and, likewise, it is also better to have very little within-area variability among row proportions because this leads to accurate estimates of fractions. Nevertheless, Romero and Pavía (2021) advocate studying the behavior of both algorithms when the number of units ob-

Table 2. *Summary of some features of the datasets used to evaluate the algorithms.*

			Average number of					
Country	Year	Elections (datasets)	voting units (min-max)	voters by units[1] (min-max)	parties (min-max)	candidates (min-max)	large p_{jk} fractions[2]	% voters in large p_{jk}[3]
NZ	2002	69	$83.2_{(30-651)}$	$554.6_{(24.5-1075.5)}$	$7.0_{(5-8)}$	$5.7_{(5-8)}$	$1.2_{(0-2)}$	$36.0_{(0.0-65.3)}$
NZ	2005	69	$81.8_{(29-698)}$	$634.5_{(28.3-1194.0)}$	$5.2_{(4-7)}$	$4.5_{(3-6)}$	$1.4_{(0-2)}$	$50.5_{(0.0-77.0)}$
SCO	2007	73	$70.2_{(22-103)}$	$411.6_{(346.3-547.1)}$	$6.0_{(5-8)}$	$5.9_{(5-8)}$	$2.6_{(0-4)}$	$59.1_{(0.0-80.5)}$
NZ	2008	70	$84.1_{(32-686)}$	$614.6_{(28.7-1094.8)}$	$5.4_{(4-6)}$	$4.4_{(3-6)}$	$1.7_{(0-3)}$	$52.5_{(0.0-80.7)}$
NZ	2011	70	$85.7_{(32-644)}$	$555.0_{(27.2-1068.0)}$	$5.6_{(4-7)}$	$4.7_{(4-6)}$	$1.4_{(0-2)}$	$49.7_{(0.0-73.5)}$
NZ	2014	71	$81.2_{(31-620)}$	$617.0_{(32.6-1124.2)}$	$5.9_{(5-7)}$	$4.7_{(3-6)}$	$1.5_{(0-3)}$	$49.9_{(0.0-73.9)}$
NZ	2017	71	$101.9_{(41-705)}$	$487.7_{(33.2-1012.7)}$	$5.2_{(4-7)}$	$4.8_{(3-6)}$	$1.3_{(0-2)}$	$47.3_{(0.0-77.9)}$
Total	–	493	$84.0_{(22-705)}$	$552.2_{(24.5-1194.0)}$	$5.8_{(4-8)}$	$4.9_{(3-8)}$	$1.6_{(0-4)}$	$49.4_{(0.0-80.7)}$

Source: compiled by the authors using data from the New Zealand (NZ) electoral commission and the Scotland (SCO) Electoral Office.

[1] These averages correspond to averages of averages. First, the average number of voters per voting unit $\sum_i N_{..i}/I$ is computed for each dataset and then the average of these averages is calculated for each year.

[2] A p_{jk} is considered a large fraction when it is higher than 0.80.

[3] The percentage of voters located in cells with large fractions, $p_{jk}(> 0.80)$, in each election/dataset is computed as $100\sum_{(j,k)\in L} N_{jk.}/\sum_i N_{..i}$, where $L = \{(j,k) : p_{jk} > 0.80\}$.

served is small, since this reduces the costs of data wrangling and would help to answer the question of whether they could be used immediately after an election, a time when the results are usually available at a high level of aggregation, for a small number of units. Thus, in order to increase the number of analysis scenarios, we build new datasets by merging voting units and/or voting options (parties and candidates). On the one hand, reducing the number of units adds difficulty to the problem, by reducing the amount of information available. On the other hand, reducing the number of voting options simplifies the problem, by decreasing the number of unknowns. In general, both operations contract the across-unit variance.

We derive three new datasets from each baseline dataset by (i) reducing the number of voting units, (ii) reducing the number of cells in the tables (the number of parties and candidates), and (iii) reducing both the number of units and the number of cells. More specifically, the initial number of units of each dataset is reduced by randomly grouping the units into a random number of groups, uniformly selected between 10 and 20, and merging them. The number of parties and candidates is reduced by adding to either the row or column voting option Others, respectively, the votes of those parties or candidates that did not reach a minimum of 20% of the votes. The random merging of units in scenarios (i) and (iii) have been performed in an independent fashion in order to induce more variability in the constructed database. After all these operations, we are ready to analyze real data equivalent to the 1972 elections.

4. An initial comparison of procedures

This section, focused on accuracy, assesses the solutions achieved after applying `ei.MD.bayes` (with different specifications), `lphom`, `tslphom` and `nslphom` to the reference database of the 493 datasets that made up our collected data before performing the processes of merging of units and/or cells described in the last paragraph of the previous section. As a rule, and starting point, we have considered the default options of the functions, given that these are usually the specifications most utilized by users. These simplify their decision-making processes, reduce their operational costs and favor automation. In the case of `ei.MD.bayes`, we consider three different specifications, which we label as `ei_default`, `ei_auto`, and `ei_manual`.

The `ei_default` solutions correspond to the use of `ei.MD.bayes` with default options. A solution based on MCMC, however, requires convergence to the equilibrium distribution of the Markov chains to be reliable. Unfortunately, this is not attained as a rule with our data when `ei.MD.bayes` is employed with default options. The arguments of the function, therefore, have to be tuned to generate convergent chains. The `eiPack` package also includes a function, `tuneMD`, with "an adaptive algorithm to generate tuning parameters for the MCMC algorithm implemented in `ei.MD.bayes`" (Lau et al., 2020). So, as a second specification for `ei.MD.bayes`, we implement a two-step strategy in which firstly `tuneMD` is employed with default options to afterwards apply `ei.MD.bayes` with its `tune.list` argument equal to the output generated by `tuneMD`. This does not solve the lack of convergence, however. This should not come as a surprise since, as the `ei.MD.bayes` help file advises: most problems will require a much larger thinning interval and a longer burn-in period than default.

At this point, it is clear that what is necessary is to manually customize the arguments of the `ei.MD.bayes` function. Unfortunately, trying to customize 493 scenarios is impractical and extremely time-consuming. An analyst needs to test several specifications for each election, plotting each of their outputs and diagnosing their convergence (Roy, 2020). Hence, as an intermediate, operative approach, we look for a specification that could work well in general. After picking at random three datasets of each block (year) of elections, we find that a two-step strategy in which firstly `tuneMD` is employed with arguments `ntunes = 10` and `totaldraws = 100000` and secondly `ei.MD.bayes` is employed with arguments `sample = 1000`, `thin = 100`, `burnin = 100000` and `tune.list` equal to the output generated by `tuneMD` reaches the convergence in all twenty-seven elections selected. We use this specification to model all the datasets. This guarantees that convergence is reached in the twenty-seven datasets checked and also, with really high probability, in the rest of datasets. We label the solutions attained by `ei.MD.bayes` using this specification `ei_manual`.

Once each algorithm is run, we gauge accuracies of solutions using as discrepancy measures the statistics EI and WPE, given by equations (16) and (17). These statistics account for the distances between the forecasted and observed matrices at the district level of, respectively, counts and proportions. The assessments of errors at the local level are unfeasible in this case as internal cell values of local units are not available in the collected datasets.

$$EI = 100 \cdot \frac{0.5 \sum_{jk} |N_{jk\cdot} - \hat{N}_{jk\cdot}|}{\sum_{jk} N_{jk\cdot}} \tag{16}$$

$$WPE = 100 \cdot \frac{\sum_{jk} N_{jk\cdot}\ |p_{jk} - \hat{p}_{jk}|}{\sum_{jk} N_{jk\cdot}} \tag{17}$$

The EI (error index) statistic is a classical measure of discrepancy (e.g., Thomsen, 1987; Klima et al., 2016; Romero et al., 2020) that quantifies the distance between matrices of counts. In our case, it accounts for the percentage of votes wrongly assigned, i.e., the minimum number of votes that should be moved among cells of the forecasted matrix to accomplish a perfect fit. Multiplication by 0.5 is done to prevent counting every incorrectly allocated vote twice. This coefficient varies between 0, when the actual and the forecasted matrices coincide, and 100, when no single vote is correctly assigned. The WPE statistic (proposed in Pavía and Romero, 2022) measures the weighted average distance between the actual p_{jk} and the estimated $\hat{p}_{jk}$ proportions, using as weights the corresponding actual counts. This statistic ponders more the discrepancies associated with the most relevant proportions and also ranges between 0, when $\mathbf{P}$ and $\hat{\mathbf{P}}$ match, and 100, when no vote is correctly assigned. EI and WPE are closely correlated.

Table 3 synthesizes the discrepancies, measured using the EI and WPE statistics, between the actual matrices and the solutions attained after applying `lphom`, `tslphom`, `nslphom` and `ei.MD.bayes` (with the three specifications detailed above) to the datasets of the reference database. The table presents, by group of elections, average figures of EI and WPE values as well as average computation times (lower panel). The upper panel of the table also offers some summary statistics of the corresponding group of elections. The elections are naturally grouped by country and year. Ultimately, all the elections of each group are related since they were held simultaneously, sharing the same general political context. The last column summarizes the results corresponding to the whole database.

Figures 1 and 2 display the same information shown in the EI and WPE panels of Table 3, but graphically. Interested readers can also consult Figure S1 of the supplementary material which displays graphically the averages times of computation (in seconds) required to reach the solutions. Several initial findings emerge analyzing Figures 1 and 2 and the numbers in Table 3. First, all the methods yield solutions superior to a random assignment. Second, as expected, the `ei_default` and `ei_auto` solutions are by far the least accurate, given their lack of convergence. They are, nevertheless, superior to a random assignment. This may seem surprising at first glance, however, despite their failure to converge, they already include the information available in the margins of the tables; an issue that limits the set of possible solutions. Third, within the `lphom` family, `nslphom` is the one that is most accurate. This confirms the conclusions reached in Pavía and Romero (2022). Fourth, both `ei_manual` and `nslphom` solutions stand out for being the most accurate, being indeed fairly good considering the magnitude of error that, according to Klima et al. (2016), is usual in these kind of problems. Fifth, the `ei_auto` and `ei_manual` specifications require much more time than the rest of the procedures to reach their solutions.

Table 3. *Summary of the performance of the algorithms in the original datasets.*

Country Year	NZ 2002	NZ 2005	SCO 2007	NZ 2008	NZ 2011	NZ 2014	NZ 2017	NZ + SCO
# of Elections	N = 69	N = 69	N = 73	N = 70	N = 70	N = 71	N = 71	N = 493
Avg. # of units	$\bar{I}$ = 83.2	$\bar{I}$= 81.8	$\bar{I}$= 70.2	$\bar{I}$= 84.1	$\bar{I}$= 85.7	$\bar{I}$= 81.2	$\bar{I}$= 101.9	$\bar{I}$= 84.0
Avg. # of cells	$\overline{RC}$= 39.5	$\overline{RC}$= 23.8	$\overline{RC}$= 35.2	$\overline{RC}$ = 23.4	$\overline{RC}$= 26.2	$\overline{RC}$= 27.9	$\overline{RC}$= 24.8	$\overline{RC}$= 28.7
Average of *EI* mesasures								
ei_default	22.75	27.69	48.33	31.19	29.26	32.40	34.38	32.42
ei_auto	25.20	28.96	46.85	30.89	30.17	33.18	33.93	32.85
ei_manual	10.75	8.53	23.09	8.34	7.68	7.88	6.93	10.52
nslphom	12.79	9.68	8.86	9.11	9.46	9.69	8.91	9.77
tslphom	14.80	11.09	11.00	10.88	11.50	11.66	10.91	11.68
lphom	16.88	12.29	12.92	12.22	12.99	12.95	12.20	13.20
Average of *WPE* mesasures								
ei_default	16.29	21.70	41.55	25.26	23.30	26.32	28.11	26.20
ei_auto	18.44	22.70	40.46	25.04	23.94	26.78	27.67	26.54
ei_manual	6.30	5.61	18.47	5.86	4.88	4.86	4.54	7.28
nslphom	7.90	6.09	4.80	6.09	6.26	6.55	5.67	6.18
tslphom	9.42	7.52	6.72	7.90	8.05	8.15	7.46	7.89
lphom	10.82	8.46	8.07	8.89	9.13	9.04	8.39	8.96
Average of computational burden (in secs)								
ei_default	2.08	1.23	1.33	1.14	1.55	1.48	1.52	1.48
ei_auto	958.57	573.53	603.36	531.03	724.65	692.13	722.93	690.06
ei_manual	1150.58	687.37	765.20	636.40	864.02	827.75	853.43	825.70
nslphom	5.41	5.32	5.88	5.85	5.61	5.28	6.80	5.74
tslphom	0.92	0.85	0.81	0.87	0.87	0.81	0.97	0.88
lphom	0.56	0.64	0.25	0.52	0.64	0.60	0.64	0.55

Source: compiled by the authors after applying the function `ei.MD.bayes` of the R package `eiPack` (Lau et al., 2020) with different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021) to the official data from the New Zealand electoral commission and the Scotland Electoral Office described in Section 3. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as `tune.list` argument and default options for the rest of its arguments and (iii) `sample = 1000`, `thin = 100`, `burnin = 100000` and the output of function `tuneMD` with `ntunes = 10` and `totaldraws = 100000` as `tune.list` argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core ™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

Looking at the outcomes of Table 3 in more detail reveals further findings. Sixth, as a rule, the performance of all methods worsen when either the number of cells grows or when the number of units decreases, but it seems that the accuracy of the `ei.MD.bayes`-based solutions suffer significantly more than the `lphom`-based solutions when the number of units decreases. Seventh, it seems that most of the time `ei_manual` produces slightly better solutions than `nslphom`, but with `nslphom` being more robust. Indeed, `nslphom` beats `ei_manual` after pooling all the elections. This, however, is a consequence of the poor performance of `ei_manual` in some Scottish datasets (see Figure S8 in the Supplementary Material) due to a lack of convergence, which in this case can be solved working with larger chains. We investigate these results further in the sections that follow.

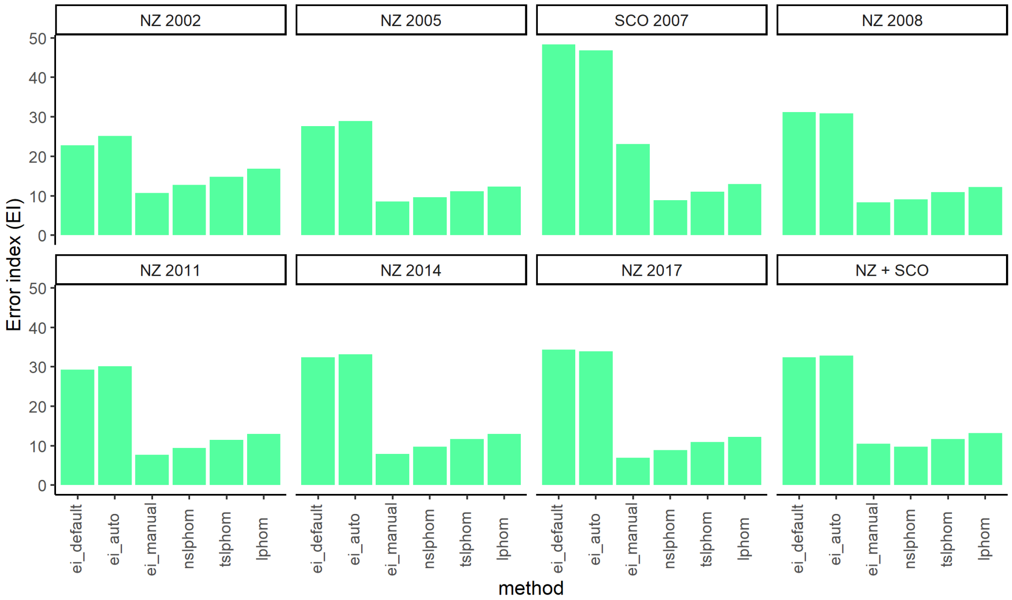

Figure 1. *Graphical representation of average values of EI error measures grouped by election and algorithm in the reference database. Individual solutions have been attained with the function `ei.MD.bayes` of the R package `eiPack` (Lau et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021) with default options. Details of the specifications used when applying `ei.MD.bayes` can be consulted at the bottom of Table 3.*

From the above list of findings, we can gain some interesting insights. Firstly, the solutions reached using the default options of `ei.MD.bayes` are, as a rule, scarcely accurate. Despite the advantages users may find in employing functions with default options without more inquiries, this should be avoided in the case of `ei.MD.bayes`. Secondly, the default solutions of `ei.MD.bayes` can be significantly improved with some extra work by tuning all its parameters. Thirdly, the functions of the `lphom` package produce highly competitive solutions in an automatic way. Finally, the `lphom`-based solutions are, at least in these examples, reached in very few seconds.

5. Sensitivity and robustness. The effects of reducing the number of units and/or cells

The previous section evaluates `ei.MD.bayes` and `nslphom` in a set of scenarios where the relationship between the amount of information available (number of units) and the complexity of the problem (number of cells in the matrix) is considered adequate. On average, there are 2.95 voting units for each parameter to estimate when, according to Plescia and De Sio (2018, p. 673), "the literature specifies a criterion of at least two [sub]units per coefficient" for a proper forecasting of district level fractions. Although the average number of cells that we have had to estimate per election is high

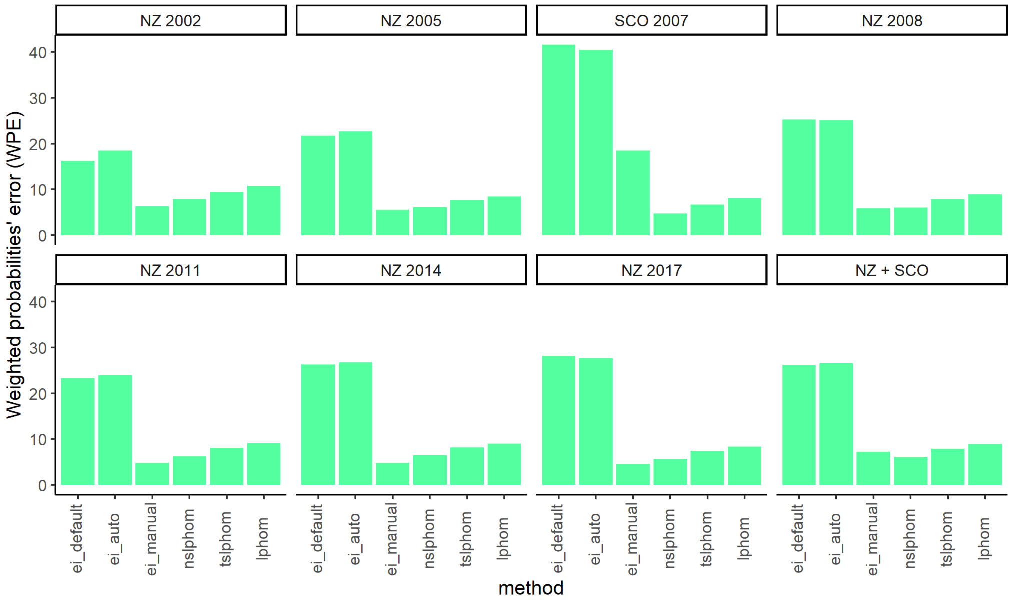

Figure 2. *Graphical representation of average values of WPE error measures grouped by election and algorithm in the reference database. Individual solutions have been attained with the function `ei.MD.bayes` of the R package `eiPack` (Lau et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021) with default options. Details of the specifications used when applying `ei.MD.bayes` can be consulted at the bottom of Table 3.*

(28.4), so is the average number of voting units available (84), with a range that varies between a minimum of 22 and a maximum of 705, although with only 6 and 36 elections above 600 units and 200 units, respectively. Under these conditions, we get, on average, predictions of a high and similar quality, both using the `ei_manual` specification of `ei.MD.bayes` and the default options of `nslphom`. In this section, we study how the different algorithms respond when adding to the problem, by reducing the number of units, and/or through its simplification, by reducing the number of unknowns.

It is important to understand the sensitivity and robustness of the estimates when using a decreased number of units because, firstly, there are situations where obtaining more disaggregated data may be limited or even impossible (for example, in historical elections) and, secondly, because, depending on its costs in terms of accuracy, it is an option worth considering as decreasing the number of units can lead to a drastic reduction in the expenses of obtaining and handling data. It is also relevant to study how the methods behave when the number of unknowns is reduced, focusing on just the main cells. After all, the analyst, on occasions, is not interested in an overall vision of the matrix but rather in certain relevant fractions/transfers.

To answer the previous research questions, we use the three new databases derived, as stated in Section 3, from the reference database. Note that we have created three additional databases, each one also composed of 493 datasets, by just (i) grouping units in each dataset, (ii) reducing (by aggregation) the number of cells to estimate in each dataset, and (iii) merging both, units and cells, in each dataset. In this section, we first analyze the impact of reducing the number of units, then we study the effect of reducing the number of cells and, finally, we examine the joint effect of both operations.

5.1. Effects of reducing the number of units

As in Table 3, Table S1 in the supplementary material summarizes the discrepancies measured using the *EI* and *WPE* statistics between the real matrices and the solutions attained after applying `ei.MD.bayes` (with the three specifications considered), `lphom`, `tslphom` and `nslphom` to the datasets obtained by randomly merging the observed units. Figure 3 and Figures S2 and S3 in the supplementary material present graphically the information of the different panels of Table S1. Given that the general picture drawn by *EI* and *WPE* measures is quite similar, the graphical representations corresponding to the *WPE* measures from Table S1, and the equivalent analysis in next two subsections are presented only in the supplementary material in order not to overburden this presentation.

Comparing the results of Tables 3 and S1 (Figures 1 and 3) it can be seen that, as expected, the accuracy of the solutions deteriorates as a consequence of the drastic reduction in the number of units. The impact, however, is not homogeneous in all methods. Reducing the number of units changes the order of preference between the algorithms. The solution associated with the `ei_manual` of `ei.MD.bayes` is the one that suffers the most. The mean error of this approximation is multiplied by more than two: `ei_manual` goes from having the lowest mean values for *EI* and *WPE* in almost all the election blocks to registering, in all cases, values clearly higher than those of all the solutions of the `lphom` family. Within this subset of solutions, however, the order is maintained, with the `nslphom` solutions clearly dominating those of `tslphom` and `lphom`, and this despite the fact that their relative deterioration within the subgroup is higher, with a mean increase in the error of 36%.

These findings are in line with Romero et al. (2020) and Romero and Pavía (2021) who, based on the study of the French presidential elections of 2017, noted that `ei.MD.bayes` suffers significantly when the number of units is reduced. Along the same lines, despite our best efforts, we have not found any general tuning of the parameters for `ei.MD.bayes` that works well with so few units. For example, the accuracy of the estimates does not improve even after multiplying the length of the MCMC chains by ten (with the configuration `sample = 10000`, `thin = 100` and `burnin = 1000000`). This is in contrast to the results of `nlsphom` which, with its default options, continues to generate fairly accurate solutions even in these scenarios. In light of these results, we can say that the `ei.MD.bayes`-based solutions are quite sensitive to the number of available units, quickly reducing their performance as soon as the number of units decreases and that, on the contrary, the `lphom`-based solutions are more robust. In terms of computing time, all solutions are achieved in fewer seconds.

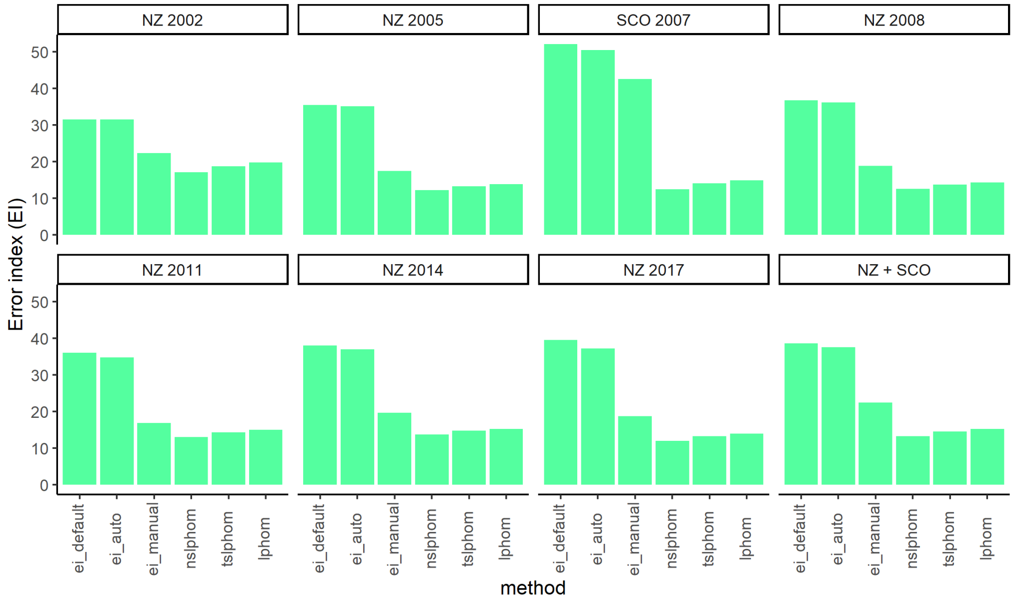

Figure 3. *Graphical representation of average values of EI error measures grouped by election and algorithm in the scenarios attained after randomly merging polling units as described in Section 3. Individual solutions have been attained with the function* `ei.MD.bayes` *of the R package* `eiPack` *(Lau et al., 2020) using three different specifications and the functions* `lphom`*,* `tslphom` *and* `nslphom` *of the R package* `lphom` *(Pavía and Romero, 2021) with default options. Details of the specifications used when applying* `ei.MD.bayes` *can be consulted at the bottom of Table 3.*

A possible explanation for the relatively worse performance of `ei.MD.bayes` in these split-ticket scenarios comes from the difficulties that its underlying (two-step) algorithm would find to move sufficiently, with so few units, the a priori row-cell fractions implied by the default values for the hyperparameters. With default options, the expected values for α_{jk} are constant by row and the expected row-cell fractions constant at $1/C$; when vote transfer matrices are characterized by having a relative large number of internal cell probabilities close to zero or one, larger than in other settings such as in racial voting applications. According to this explanation, `ei.MD.bayes` should suffer less in situations with fewer extreme fractions and/or with a lesser proportion of voters in cells with high p_{jk}. The likelihood of this explanation grows when (i) one relates the average accuracies attained in Scottish and NZ elections and their relative numbers of rows with a p_{jk} close to one (higher than 0.80) –44.1% of rows in Scotland tables and 24.3% of rows in NZ tables have a proportion close to one– or after (ii) observing no impact in the accuracy of `ei.MD.bayes` solutions when the number of units in the `senc` dataset available in the `eiPack` package is reduced. In the `senc` dataset only 26% of voters are located in cells where $p_{jk} > 0.80$. It should be noted that with this dataset of racial voting `nslphom` neither suffers a decrease of accuracy after a reduction in the number of units.

5.2. Effects of reducing the number of cells

Table S2 in the supplementary material measures, using EI and WPE, the accuracy of the solutions achieved after running `ei.MD.bayes` (with the three specifications considered), `lphom`, `tslphom` and `nslphom` in the datasets obtained by aggregating in Others the election options not surpassing 20% of the vote. Figure 4 and Figures S3 and S4 in the supplementary material depict graphically the information of the different panels of the table.

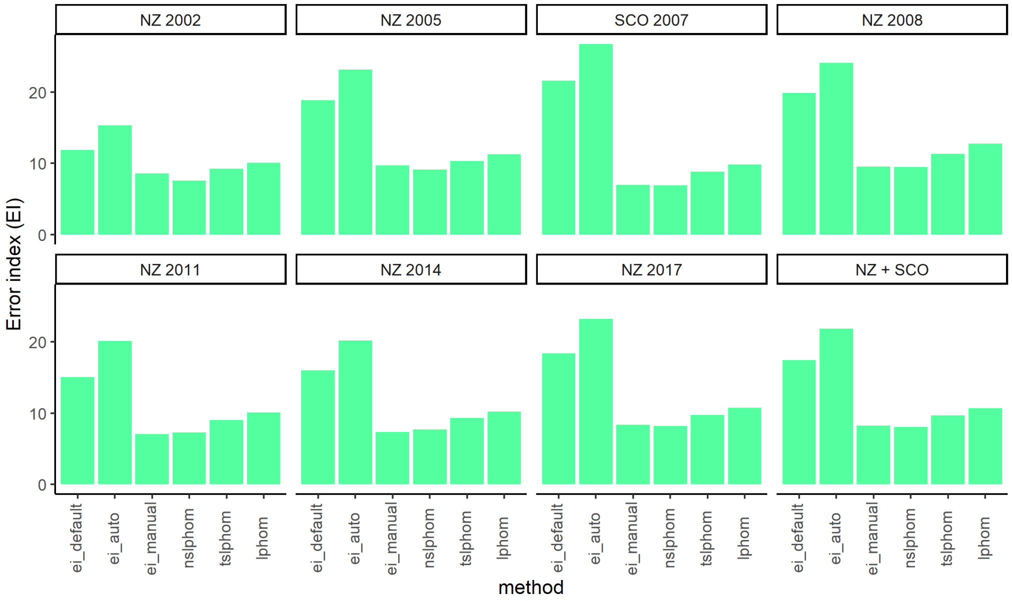

Figure 4. *Graphical representation of average values of EI error measures grouped by election and algorithm in the scenarios attained after merging in Others the election options not surpassing 20% of the vote. Individual solutions have been attained with the function* `ei.MD.bayes` *of the R package* `eiPack` *(Lau et al., 2020) using three different specifications and the functions* `lphom`, `tslphom` *and* `nslphom` *of the R package* `lphom` *(Pavía and Romero, 2021) with default options. Details of the specifications used when applying* `ei.MD.bayes` *can be consulted at the bottom of Table 3.*

Comparing the results of Tables 3 and S2 (Figures 1 and 4), it can be seen that, as expected, the accuracy of the solutions improves as a consequence of the reduction in the number of unknowns (number of cells in the transfer matrices). The general situation with respect to the baseline scenario does not change substantially. The `ei_default` and `ei_auto` specifications still do not converge, despite the reduction of unknowns, and continue to be the ones with the worst performance, while `ei_manual` and `nslphom` are the ones with the best figures, with `lphom` and `tslphom` generating highly competitive solutions. Particularly noteworthy is the fact that now the solutions for the Scottish elections with the specification `ei_manual` from `ei.MD.bayes` are significantly improved, as now all of them reach convergence. This fact means that in aggregate terms `ei_manual` is the one that most reduces its joint mean error in these scenarios (the mean of *EI* goes from 10.52 to 8.22, a reduction of almost 22%). However, taking the Scottish results out of the equation, among the two main algorithms (`ei_manual` and `nslphom`), `nslphom` is revealed as the one that benefits most from the simplification of the problem. On average, it happens to be the most accurate in five of the seven election groups, when in the reference database it was only the most accurate in one of the election groups. The relative increase of rows in the target tables with a cell where $p_{jk} > 0.80$ plays, as previously discussed, against `ei.MD.bayes` as a consequence of the a priori row-cell fractions implied by the default priors for the hyperparameters. In terms of computing times, logically, costs are reduced.

5.3. Interaction effects. Effects of reducing both the number of units and cells

In subsection 5.1, we studied the effect of having fewer units and we found that solutions based on `ei.MD.bayes` suffer markedly when the number of units decreases. In subsection 5.2, we analyzed the impact of working with problems with fewer unknowns and we found that all algorithms improved their performance. In this subsection, we study what happens when both situations occur simultaneously. Table S3 in the supplementary material presents, using *EI* and *WPE*, the accuracy of the solutions reached with `ei.MD.bayes` (with the three specifications considered), `lphom`, `tslphom` and `nslphom` in the datasets obtained after reducing the number of cells and units, as stated in Section 3. Figure 5 and Figures S5 and S6 in the supplementary material show graphically the information of the different panels of Table S3.

Comparing the results of Tables 3 and S1 to S3, and the corresponding graphical representations (Figures 1 to 5), it can be seen that in this scenario the accuracies of the solutions generated by the different algorithms are at some intermediate point between the accuracies of the solutions obtained in the analyzed scenarios in subsections 5.1 and 5.2. The relative impact of both types of reductions (of data and of unknowns), however, is not homogeneous in all algorithms, at least for these datasets and with the reductions in the number of units and cells implemented. In the case of solutions based on `ei.MD.bayes`, we see that reducing the number of units has much more impact than reducing the number of cells, while in the case of solutions based on `lphom` the opposite relationship is observed, with the decrease in the number of unknowns having more relative importance. These results confirm and reinforce the conclusions reached in the previous subsections: `ei.MD.bayes` inferences are very sensitive to the data-unknowns relationship, deteriorating notably when the level of detail of the information is reduced, while `nslphom` is very robust, being more insensitive to a decrease in the amount of available data. In all cases, computing times very clearly drop.

6. A comparison of ei_manual and nslphom solutions

From the analyses carried out in sections 4 and 5, we can affirm that the `ei_manual` and `nslphom` algorithms are clearly the ones that provide, within each methodology

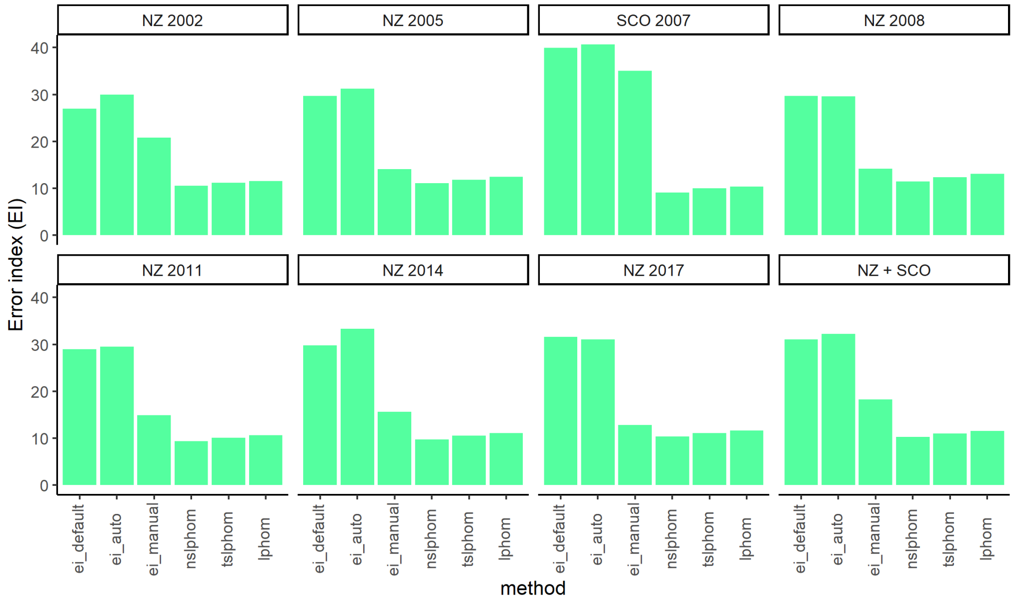

Figure 5. *Graphical representation of average values of EI error measures grouped by election and algorithm in the scenarios attained after merging in Others the election options not surpassing 20% of the vote and randomly merging polling units as described in Section 3. Individual solutions have been attained with the function* `ei.MD.bayes` *of the R package* `eiPack` *(Lau et al., 2020) using three different specifications and the functions* `lphom`*,* `tslphom` *and* `nslphom` *of the R package* `lphom` *(Pavía and Romero, 2021) with default options. Details of the specifications used when applying* `ei.MD.bayes` *can be consulted at the bottom of Table 3.*

(model statistical approach and mathematical programming), the most accurate solutions in our databases. The behavior of both sets of solutions, however, is not homogeneous, presenting important variations among datasets within and between algorithms. In fact, as can be seen in Figure 6, which displays graphically a summary of the average values of EI and WPE in each database for the `ei_manual` and `nslphom` solutions, although `ei_manual` and `nslphom` present (on average) predictions of equivalent quality when the number of available units is large enough, both start to differ clearly when the amount of available data decreases, with the `ei_manual` solutions deteriorating faster. In this section, we look at the analysis in more detail. Focusing exclusively on these two procedures, we investigate, on the one hand, the factors that influence their global accuracies and their differences in accuracy and, on the other hand, the characteristics of the estimates obtained by both algorithms for the fractions p_{jk}. The insights extracted from these latter analyses might open a way forward for exploring how to improve a forecast by combining solutions.

Specifically, after analyzing the distributions of EI and WPE values obtained by both procedures in the entire set of datasets, we investigate the relationship between the accuracies obtained and some of the main characteristics associated with each dataset. With this, we aim to determine what the relative impact of each feature is and to understand

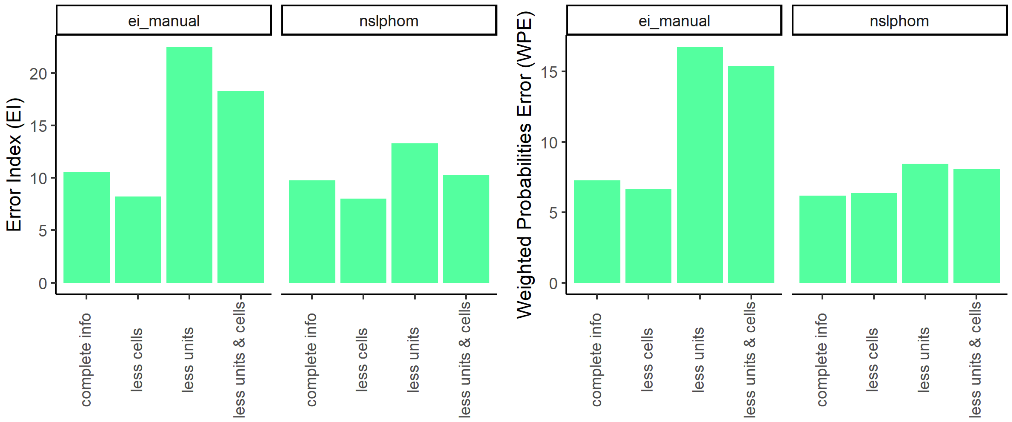

Figure 6. *Graphical representation of global average values of EI and WPE error measures grouped by database for* `ei_manual` *and* `nslphom`*. Individual solutions have been attained with the function* `nslphom` *of the R package* `lphom` *(Pavía and Romero, 2021) with default options and the function* `ei.MD.bayes` *of the R package eiPack (Lau et al., 2020) with customized options. Details of the specification used when applying* `ei.MD.bayes` *can be consulted at the bottom of Table 3.*

under what circumstances each of the methods could work better. This study, focused on the analysis of global accuracies, is complemented by a more detailed look at the cells of the matrices. The second subsection of this section is dedicated to analyzing the quality and properties of the estimates of the fractions p_{jk} that are obtained with both procedures. The analysis is relevant because, according to some authors (e.g., Upton, 1978; Johnston and Hay, 1983), the methods based on mathematical programming have a tendency to predict extreme fractions; the opposite bias attributed by Romero and Pavía (2021) to `ei.MD.bayes`. In the last subsection, we take advantage of these insights to propose a simple rule that can be used to improve forecasts in certain circumstances.

6.1. Factors impacting on the accuracy of the procedures

Figure 6 suggests the existence of important differences in terms of accuracy in the solutions generated by `ei_manual` and `nslphom` and that these depend on the characteristics of the electoral processes under study. Figure 7, where the distributions obtained for *EI* and *WPE* with both procedures are plotted in the 1972 datasets analyzed, clearly shows the existing variability in the solutions reached by each method and between methods (in Table S4 of the supplementary material the interested reader can consult a statistical summary of both distributions). For example, focusing on *EI* (the conclusions for *WPE* would be very similar, see Table S4), we observe that the errors associated with `nslphom` are, on average, more than 4 points lower than those of `ei_manual`. Another interesting observation is that `ei_manual` errors are significantly more dispersed than those obtained by `nslphom`, with respective standard deviations of 10.8 and 4.6. Both results confirm a fact already discussed above: `nslphom` is in this database not only somewhat better on average but it is also more robust. In fact, although the distance between the medians is much lower than that observed for the means, just 0.76, it continues to be statistically significant, with a p-value smaller than 0.000001 in the sign test for paired data.

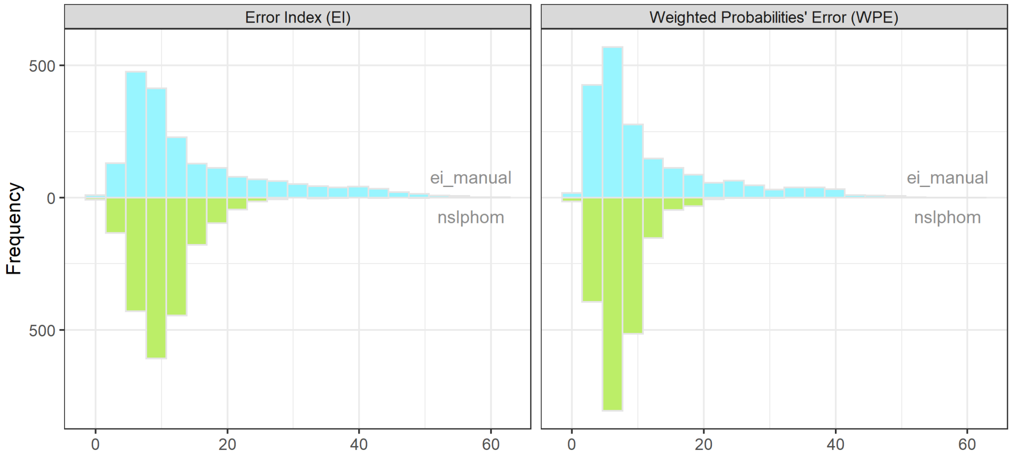

Figure 7. *Histograms of the distributions of the error measures (EI left panel and WPE right panel) linked to the solutions attained after running* `nslphom` *with default options and* `ei.MD.bayes` *with the* `ei_manual` *specification (see the bottom of Table 3 for details) in the 1972 datasets analysed in this research (see Section 3 for details).*

Figure 7 (and Table S4) clearly show that there is a high variability in the accuracies of the results obtained, so it is worth asking what the factors are that would explain, at least in part, the high variability observed within and between methods. Using multiple regression models with EI and WPE as response variables, in this subsection we study the impact that some of the main variables that characterize the scenarios considered have on accuracy. Given the great diversity we have (part of which can be seen in Table 2), we consider this analysis will give us general results regarding the behavior of the two methods rather than about idiosyncrasies of the particular data analyzed, although this cannot be completely discarded.

In addition to variables already considered throughout this paper related to the amount of information available, I, or the complexity associated with the problem, JK, other factors, such as the variability or the degree of dependence presented by the data, have also been proposed in the literature as determinants of the quality of the estimates. Table 4 details the variables considered. Table S5 in the supplementary material presents a statistical summary of the values obtained for the nine variables introduced in Table 4 in the 1972 datasets analyzed, and Table S6 offers the corresponding correlation matrix. A high correlation (0.86) can be seen between both measures of across unit variances on the patterns of votes, *var.Part* and *var.Cand*, with the correlation between $std.Part$ and $std.Cand$ also being high (0.62). In any case, given the large sample size, we do not expect this to pose a problem in interpreting the models obtained.

Table 4. *Features considered in the models.*

Variable	Description
I	Number of units. Indicator of the quantity of information.
JK	$J \times K$, number of cells in the matrix. Indicator of the complexity of the problem.
JKratio	Quotient J/K. This captures the impact of the asymmetric role played by the two dimensions of the transfer matrix. The algorithms estimate the parameters of J (multinomial) distributions, each one of dimension $K-1$.
HET	Actual heterogeneity index. This measures the degree of non-compliance of the homogeneity hypothesis: $HET = 50(\sum_{ki} \lvert \sum_j N_{j \cdot i} p_{jk} - N_{\cdot ki} \rvert / \sum_{ij} N_{j \cdot i}$. Although this coefficient cannot be computed in regular applications (as the transfer matrix is unknown), it may be estimated.
$Chi2$	Standardized χ^2-Pearson statistic of independence of the global matrix of counts. This measures the degree of dependence between the row and column categories: $Chi2 = \sum_{jk}(N_{jk\cdot} - N_{\cdot k \cdot} N_{j \cdot \cdot})^2 / [(J-1)(K-1)\sum_{jk}(N_{\cdot k \cdot} N_{j \cdot \cdot})]$. Although this coefficient cannot be computed in regular applications, it may be estimated.
$var.Part$	Compositional total variance (Pawlowsky-Glahn, Egozcue and Tolosana-Delgado, 2015) of the marginal row distributions in the I units. This measures to what extent party vote supports are different across units: $(2J)^{-1} \sum_{j,j'}^{J} Var(\{log(N_{j \cdot i}/(N_{j' \cdot i})\}_i)$.
$var.Cand$	Compositional total variance of the marginal column distributions in the I units. This measures to what extent candidacies vote supports are different across units: $(2J)^{-1} \sum_{k,k'}^{K} Var(\{log(N_{\cdot ki}/(N_{\cdot k'i})\}_i)$.
$std.Part$	Standard deviation of the distribution of percentages of votes to parties in the whole electoral space. Indicator of the degree of vote concentration/variability among parties: $sd(\{N_{j \cdot \cdot}/N_{\cdot \cdot \cdot}\}_j)$
$std.Cand$	Standard deviation of the distribution of percentages of votes to candidacies in the whole electoral space. Indicator of the degree of vote concentration/variability among candidacies: $sd(\{N_{\cdot k \cdot}/N_{\cdot \cdot \cdot}\}_k)$

Source: compiled by the authors.

In order to facilitate the interpretation of the parameters of the fitted models, all the explanatory variables have been standardized, to zero mean and standard deviation 1. In this way, the relative importance of each variable can be directly assessed as it is proportional to the value estimated for its coefficient in the regression model. Approximately, this value multiplied by four quantifies the expected variation in the response variable due to the fluctuation in the sample of the variable considered. Table 5 shows the coefficients of the fitted models. Tables S7 to S12 in the supplementary material show the obtained models in more detail.

Table 5. *Impact of different electoral features on ecological inference solutions' accuracy.*

Variable	Response variable: *EI*			Response variable: *WPE*		
	`nslphom`	`ei_manual`	difference	`nslphom`	`ei_manual`	difference
Constant	10.2110^{***}	14.5998^{***}	4.3888^{***}	7.2631^{***}	11.3220^{***}	4.0589^{***}
I	-1.2926^{***}	-1.9574^{***}	-0.6648^{**}	-0.9977^{***}	-1.5012^{***}	-0.5035^{**}
JK	1.5147^{***}	2.9153^{***}	1.40066^{*}	-0.6129^{**}	-0.0235	0.5894
*JK*ratio	0.7677^{**}	0.6548	-0.1129	1.5018^{***}	2.0344^{***}	0.5326
HET	2.4321^{***}	0.9372^{***}	-1.4949^{***}	1.7783^{***}	0.1746	-1.6037^{***}
*Chi*2	-0.7034^{***}	-1.2736^{***}	-0.5702^{*}	-0.1495	-0.6517^{**}	-0.5022^{*}
var.Part	0.4160^{*}	-3.0448^{***}	-2.6288^{***}	-0.2221	-2.6846^{***}	-2.4625^{***}
var.Cand	-1.8088^{***}	-1.1346^{**}	0.6742	-1.2098^{***}	-0.2001	1.0097^{**}
std.Part	0.4046^{***}	-1.8130^{***}	-2.2176^{***}	0.2740^{***}	-1.8953^{***}	-2.1693^{***}
std.Cand	-0.6001^{***}	-2.5661^{***}	-1.9660^{***}	-0.3760^{***}	-2.1921^{***}	-1.8160^{***}
Adjusted $R^2(\%)$	42.48	30.97	21.44	28.76	26.52	21.65
Std resid. error	3.49	8.97	9.29	2.89	8.16	8.43

Source: compiled by the authors. All the predictor variables were standardized before fitting the models to make comparisons of coefficients easier. ***, p-value < 0.01; **, p-value < 0.05; *, p-value < 0.10. More details of the fitted models can be consulted in Tables S7 to S12 in the supplementary material.

A total of six models were adjusted in order to identify the variables that impact on the quality of predictions (see Table 5). For each discrepancy measure (*EI* and *WPE*) and also for their differences, we adjusted a model to the errors obtained with each of the algorithms (`nslphom` and `ei_manual`). We now focus on analyzing the results obtained for the models using *EI* as the response variable, since the interpretations with *WPE* are similar.

Of the nine variables considered and taking as reference a p-value smaller than 0.01, seven would be selected when analyzing the errors that `nslphom` makes (see the first column of estimates in Table 5). All variables, except *JKratio* and *var.Part*, show a statistically significant impact (p-value < 0.01). Together these variables explain 42% of the observed variability. The complexity of the problem (*JK*), its heterogeneity (*HET*) and the variability across units of the target marginal distributions (*var.Cand*) are revealed as the variables with the greatest effect. Specifically, as expected, the error grows as the complexity of the problem increases and there is greater heterogeneity. Likewise, the errors decrease when there is greater variability in the marginal target distributions. Along with these variables, the amount of information available (*I*), the standard deviations of the global distributions of parties and candidates (*std.Part* and *std.Cand*) and the degree of dependency (*Chi*2) between parties and candidates are also significant. Of these variables, the amount of information is the one that has the greatest impact, and with the expected sign. The error grows as the amount of information available decreases.

The next column offers the adjusted model when analyzing the errors associated with the predictions obtained with `ei_manual`. On this occasion, the model has less explanatory power. However, the same variables identified in the previous model are maintained, and with the same signs. Using 0.01 as cutoff for significance, the main change lies in the inclusion of the variable *var.Part*, which measures the variability in the marginal distributions of origin. This result is in line with Wakefield (2004), who also in a Bayesian framework states that having smaller within-area variability among row proportions leads to more accurate estimates of fractions. As a rule, it can be seen that in both models the error grows with the complexity of the problem, when the amount of information available decreases or when there is more heterogeneity (i.e., there is more variability between units in the transfer matrices), while the error decreases when there is a greater variety in the data (variance across units) and when there is a greater relationship between the options of the rows and columns. All these variables had already been identified, in one way or another, as determinants for the quality of the estimates (e.g., King, 1997; Park et al., 2014; Klima et al., 2016; Plescia and De Sio, 2018). The relative importance of each of them, however, varies for both methods. It is worth highlighting the fact that the variable *var.Part* which measures the diversity in the marginal distributions of origin, previously identified as a key in the (Bayesian) ecological inference literature, does not appear as a determinant for `nslphom`, where it is subsumed by the variable *var.Cand* which measures the diversity in the marginal target distributions.

In comparative terms, and focusing now on the analysis of the differences (see third column of estimates in Table 5), we can see that although the impact of the amount of information available (*I*) and of the variability across units in the target distributions (*var.Cand*) affects both methods in a similar way, other variables such as heterogeneity or complexity of the problem do not. The `nslphom` algorithm is more sensitive to non-fulfilment of the homogeneity hypothesis on which it is based, while, in contrast, the `ei_manual` suffers more when the complexity of the problem increases. Likewise, although both methods depend on the variability between the marginal distributions of the territorial units (note that if *var.Cand* or *var.Part* were null, neither of them would be able to reach a solution), `ei_manual` has a greater dependence on *var.Part*, the variability across units between the row marginal distributions. The rest of the variables also have a greater impact on the quality of the `ei_manual` estimates; their estimates improve relatively when there are more differences in sizes between origin and destination options and a greater degree of dependence between them.

Finally, in order to study the possible non-linearity of the effects of the different variables, we also estimate new models in which we consider, in addition to the variables detailed in Table 4, their squares as predictors. The results of these new models, available in Tables S13 to S15 of the supplementary material, reveal the existence of significant quadratic effects for almost all the variables considered; the signs of the curvatures being contrary to those observed for the corresponding linear effect. The conclusion from this is that the estimated effects on *EI* of an increase in value of the different explanatory variables are especially acute for low values, but diminish as the values increase.

6.2. An analysis of the errors in the estimation of p_{jk}

Once the global adjustments of the matrix forecasts have been analyzed in depth, we focus on the individual cell estimates. In the reference set of 493 elections, a total of

14158 proportions, p_{jk}, were estimated. The results associated with the datasets obtained by random merging of units and/or election options are not considered in this analysis since the collapses do not modify the actual p_{jk} values. The left and middle panels of Figure 8 show the histograms, real and estimated, for `ei_manual` and `nslphom` of the 14158 p_{jk} coefficients. The histograms are found to be slightly bimodal, with a marked accumulation of frequencies in the low values, a continuous decrease as the value of p_{jk} increases and a slight rebound for the highest values.

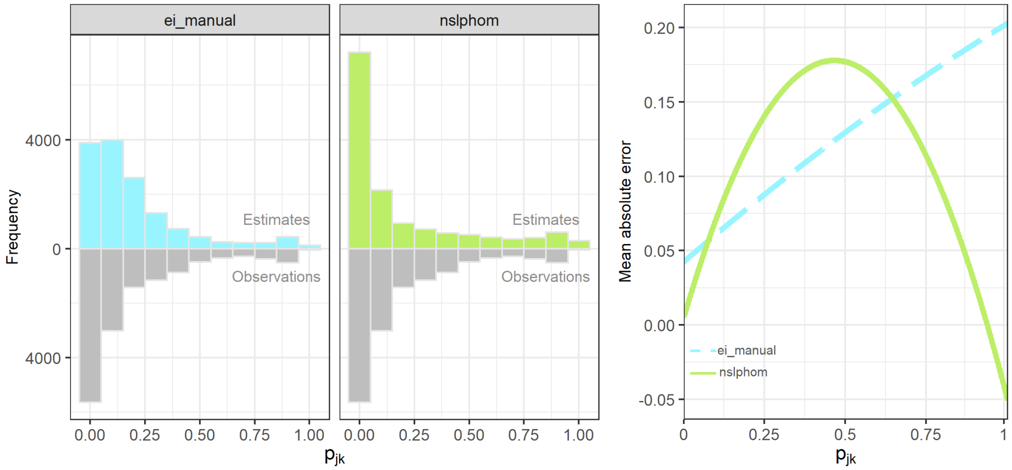

Figure 8. *Histograms of the distributions of* `ei_manual` *(left panel) and* `nslphom` *(centre panel) estimates for p_{jk} and stylized relationships between mean absolute errors of estimates and actual values (right panel). To make the comparisons easier, left and centre panels also offer the actual distributions of the p_{jk} proportions. The displayed* `ei_manual` *and* `nslphom` *estimates correspond to the solutions attained after applying* `ei.MD.bayes` *using the* `ei_manual` *specification and* `nslphom` *with default options to the 493 datasets of the reference database. The curve relationships of the right panel have been obtained after fitting the absolute value errors of the forecasts, $|p_{jk} - \hat{p}_{jk}|$, as a quadratic function of the p_{jk} proportions.*

These forms are a logical consequence of the fact that in split-ticket electoral contexts there are usually close links between the column and row options (in our examples, between parties and candidates), which give rise to the presence of values close to 1 in some rows of the probability matrix (see Table 2), chiefly in the party-rows of the leader candidates. A value close to 1 in a row necessarily implies $C-1$ values close to 0 in that same row. As can be seen in the histograms (see Figure 8), there are numerous values close to 0 and a smaller but relevant number of values relatively close to 1, with still a relevant presence of intermediate values. Intermediate values tend to be more abundant, however, in demographic voting.

The left and center panels of Figure 8 show that the biases attributed in the literature (Upton, 1978; Johnston and Hay, 1983; Romero and Pavía, 2021) to methods based on mathematical programming and `ei.MD.bayes` are manifested in our application: `nslphom` tends to estimate a higher percentage of extreme values and `ei.MD.bayes` to underestimate them. This fact is also reflected in a bias analysis. Table 6 shows the mean values of the errors of both procedures in the estimation of the p_{jk}, differentiated according to the fact that whether the real values are less than 0.20, greater than 0.80 or intermediate between both limits. On average, the biases are significantly higher for the `ei_manual` than for `nslphom` (see third and fourth columns of Table 6), with a different behavior in both procedures. While `nslphom` tends to overestimate high p_{jk} values and underestimate low values, `ei_manual` tends to overestimate low values and underestimate high values.

Table 6. *Average biases and mean absolute errors (MAE) grouped by intervals of p_{jk}.*

	Number of	Average bias (×100)		Average MAE (×100)	
Range	observations	`nslphom`	`ei_manual`	`nslphom`	`ei_manual`
$0.0 \leq p_{jk} < 0.2$	9407	−0.46	4.04	4.27	5.42
$0.2 \leq p_{jk} < 0.8$	3973	0.64	−6.85	15.27	11.29
$0.8 \leq p_{jk} \leq 1.0$	778	2.37	−13.90	4.29	17.78

Source: compiled by the authors.

The problem with calculating mean biases is that they do not reflect the true magnitude of the errors, as they include individual biases with opposite signs in their calculation. To correct this issue, the last two columns of Table 6 provide the mean values of the errors in absolute values. From these data it is clear that `nslphom` is somewhat more precise than `ei_manual` when estimating low values of p_{jk}, less precise when estimating intermediate values and, above all, much more precise when estimating high values. This is clearly a consequence of their default underlying algorithms: `nslphom` takes as seed the `lphom` solution which tends to favor extreme points of the convex hull of the region of feasible solutions defined by the constraints, whereas `ei.MD.bayes` starts, at the very bottom level of the hierarchy, by stating a symmetric distribution that assumes no prior differences between the fractions in each row. To more clearly visualize the situation, the absolute errors obtained by both procedures are adjusted as a function of p_{jk} using a quadratic model. The results of the adjustments are given in the right panel of Figure 8, with their equations available in Tables S16 and S17 of the supplementary material.

Figure 8 (right panel) shows that, as a rule, the estimation errors of `nslphom` are lower than those of `ei_manual` for values of p_{jk} which are lower than 0.10 and higher than 0.65. However, their errors are higher, on average, for intermediate values. The average superiority of `nslphom` over `ei_manual` in the analyzed examples is partially supported, therefore, by the fact that extreme values tend to be frequent in electoral studies of vote transfer. At the cost of automation, therefore, the analyst could reduce the expected bias committed by `ei.MD.bayes` using priors that place higher probabilities on larger fractions for the cells corresponding to intersections of options naturally related among the row and column categories, such as the party and the candidate of the party in ticket-splitting analysis or the same party in voter transition problems. In a mirror fashion, the analyst could also reduce the expected bias committed by `nslphom` in intermediate fractions by adding new constraints in the model for them. Constraints that reduce their initial space of feasible values from the whole [0,1] interval to some meaningful subinterval. The latter may be considered as reasonable in demographic voting studies.

6.3. Can estimates be improved by combining nslphom and ei_manual solutions?

The previous analyses give clues as to when ei_manual and nslphom will generate good solutions and also demonstrate that both methods show complementary biases in the estimates of the p_{jk}. This knowledge could be used to improve, on average, the predictions obtained using either of the two methods separately.

On the one hand, we now know that the solutions generated by ei.MD.bayes without customizing priors are, as a rule, not reliable when the number of observations is very low. On the other hand, the results suggest that nslphom generates robust solutions in a variety of situations. Both results would lead us to clearly recommend nslphom when the number of units for which information is available is low and, in general, when it is difficult to achieve convergence in the MCMC chains on which ei.MD.bayes is based.

In the above analyses, we have also learned the effect different characteristics of the analyzed scenario have on the aggregated errors, and have also verified that the errors and biases committed by ei_manual and nslphom are complementary. This last insight could be used to improve, combining both solutions, the individual predictions obtained by each method. We consider that the solutions of nslphom could always enter the equation and that the solutions of ei.MD.bayes should not enter if we cannot guarantee convergence in the MCMC chains associated with their solutions. Table 7 offers the result of combining (with the same weights) the solutions achieved with ei_manual and nslphom in the reference database. As can be seen, the combined solutions are, on average, more accurate than the individual solutions. The exception is the solutions that are achieved for Scotland, where the combined solutions are worse than those obtained with nslphom.

A detailed analysis of the solutions achieved for Scotland reveals, as shown in Figure S8 of the supplementary material, that the distribution of errors for the solutions achieved with ei_manual presents two populations. This is because the algorithm included in ei.MD.bayes only achieves, with the ei_manual specification, convergence in about half of the elections. In these scenarios, when ei.MD.bayes does not reach convergence, the analyst must decide between two alternatives: consider only the nslphom solution or manually tune ei.MD.bayes in each of the elections until the convergence of the chains can be guaranteed. This second alternative plays against automation and is quite time-consuming, being almost prohibitive when the number of elections to analyze is very high.

The Scottish results therefore raise an important question about when we can combine the solutions of ei.MD.bayes and nslphom. The obvious answer would be:

Table 7. *Summary of the performance of the solutions attained in the reference database by averaging* nslphom *and* ei_manual *solutions.*

Country Year	NZ 2002	NZ 2005	SCO 2007	NZ 2008	NZ 2011	NZ 2014	NZ 2017	NZ + SCO
# of Elections	N = 69	N = 69	N = 73	N = 70	N = 70	N = 71	N = 71	N = 493
Avg. # of units	$\bar{I}$ = 83.2	$\bar{I}$= 81.8	$\bar{I}$= 70.2	$\bar{I}$= 84.1	$\bar{I}$= 85.7	$\bar{I}$= 81.2	$\bar{I}$= 101.9	$\bar{I}$= 84.0
Avg. # of cells	$\overline{RC}$= 39.5	$\overline{RC}$= 23.8	$\overline{RC}$= 35.2	$\overline{RC}$ = 23.4	$\overline{RC}$= 26.2	$\overline{RC}$= 27.9	$\overline{RC}$= 24.8	$\overline{RC}$= 28.7
				Average of *EI* mesasures				
ei_manual	10.75	8.53	23.09	8.34	7.68	7.88	6.93	10.52
nslphom	12.79	9.68	8.86	9.11	9.46	9.69	8.91	9.77
combined	9.39	7.90	14.09	7.44	7.12	7.05	6.87	8.58
				Average of *WPE* mesasures				
ei_manual	6.30	5.61	18.47	5.86	4.88	4.86	4.54	7.28
nslphom	7.90	6.09	4.80	6.09	6.26	6.55	5.67	6.18
combined	5.76	5.23	10.22	5.16	4.71	4.49	4.50	5.75

Source: compiled by the authors after applying the function `nslphom` of the R package `lphom` (Pavía and Romero, 2021) with default options and the function `ei.MD.bayes` of the R package `eiPack` (Lau et al., 2020) with arguments `sample = 1000, thin = 100, burnin = 100000` and the output of function `tuneMD` with `ntunes = 10` and `totaldraws = 100000` as `tune.list` argument to the official data from the New Zealand electoral commission and the Scotland Electoral Office described in Section 3. Combined solutions have been obtained as arithmetic means of the `ei_manual` and `nslphom` solutions.

when we have reached convergence with `ei.MD.bayes`. This brings us back to the starting point: we have to check convergence (a process not easily automatable) and, if this is not achieved, we have to continue testing specifications, with their enormous associated labor and computational costs. To break this cycle, it would be interesting to study if there is a way to use the robust `nslphom` solution to determine 'automatically' when the solution reached by `ei.MD.bayes` is reliable.

7. Discussion and concluding remarks

The problem of forecasting the inner-cells counts of a contingency table just knowing its row and column aggregates outlines a relevant problem in many settings, including economics, epidemiology and marketing, being sociology and political science where it has aroused more interest. Social scientists, politicians and the media, among other agents, are very interested in mapping the transitions in preferences of voters between elections and in knowing how different social groups vote. Surveys are sometimes used to answer these questions. However, they are not always available (as in historical or local elections) and, more importantly, they are not especially reliable in estimating the coefficients p_{jk}. Polls present significant weaknesses in terms of both precision and accuracy (see, e.g., Miller, 1972; King, 1997; Klima et al., 2016; Dassonneville and Hooghe, 2017; Plescia and De Sio, 2018; Romero et al., 2020). Hence, a number of algorithms have been suggested in the literature to estimate from observed aggregate data the fractions p_{jk} and p^i_{jk}. Because aggregate data are readily available, the issue is to ascertain the performance of the different algorithms.

Several papers have focused on studying theoretically under which circumstances the forecasts obtained would be reliable and how the basic models can be modified under specific circumstances (see, e.g., Firebaugh, 1978; Gelman et al., 2001; Greiner and Quinn, 2009; Forcina and Pellegrino, 2019). The aim of this paper has been to assess, from an empirical perspective, the accuracy and efficiency, among other issues, of the two more powerful methods currently available for forecasting R×C ecological tables: on the one hand, the ecological Bayesian approach programmed in the `ei.MD.bayes` function of the `eiPack` R-package (Lau et al., 2020) and, on the other hand, the mathematical programming algorithms available in the `lphom` R-package (Pavía and Romero, 2021).

In this study, we have started from a singular database made up of almost 500 elections, where we have the gold standard for comparison: the real p_{jk} values, a quite unusual issue (Pavía, 2022). From this baseline database, we have created new scenarios of analysis to evaluate how the different algorithms behave in either more stressful or simpler situations. The results show that to obtain satisfactory solutions with `ei.MD.bayes` it is absolutely essential to properly tune its arguments. It is necessary to guarantee convergence in the MCMC chains on which the algorithm implemented in `ei.MD.bayes` is based in order to obtain reliable solutions. This requires adequately qualified analysts and is accompanied by significant time costs in terms of workforce and computational skills. In contrast, the `lphom` functions, especially the `nslphom` function, are capable of producing accurate results in seconds with their default options, which also makes it robust to claims of hacking. In any case, when `ei.MD.bayes` is properly tuned and convergence is reached (although, sometimes this is more difficult, such as when the amount of information available is scarce) its solutions tend to be slightly more accurate than those of `nslphom`.

In terms of robustness, it is obtained that while `ei.MD.bayes` solutions are much more sensitive to the different characteristics of the dataset used, `nslphom` generates satisfactory solutions in a significantly greater range of scenarios. The inferences of `ei.MD.bayes` with default priors are very sensitive to the data-unknowns relationship, deteriorating notably when the number of units is reduced and, more intensively, when the proportion of rows with extreme fractions grows, while `nslphom` is more robust, being quite insensitive to a decrease in the amount of available data.

The fact that `ei.MD.bayes` malfunctions with few units without proper customization and that `nslphom` generates satisfactory solutions even under those circumstances makes `lphom`-based approaches also preferable in terms of data wrangling. In fact, the costs of obtaining and pre-processing data are generally very relevant in actual ecological inference applications and they grow with the number of units. The `ei.MD.bayes` function also requires that $\sum_k N_{j.i} = \sum_j N_{.ki}$ be verified for all units, $\forall i$, which does not always occur naturally, it being necessary therefore to apply some data pre-processing strategies to guarantee the equalities (Klima et al., 2016). The functions in `lphom`, on the other hand, are capable of handling various scenarios with discrepancies in the previous accounting equalities (Pavía, 2023).

In view of all the previous considerations, our recommendation would be to use nslphom as a reference algorithm and to also use ei.MD.bayes when we are able to guarantee the convergence of the MCMC chains in the solution provided. In this case, it would even be a good idea to combine both solutions since the biases committed by both functions in the estimation of the coefficients p_{jk} are complementary. While nslphom tends to overestimate high p_{jk} values and underestimate low values, ei_manual tends to overestimate low values and underestimate high values. This result prompts us to tackle a new line of research to find ways to determine the weights with which the solutions of both functions should be combined to obtain more accurate joint solutions.

We have seen that the accuracy of the solutions achieved by both procedures depends on a set of variables that can be calculated a priori, from the observed data. For example, the nslphom algorithm is more sensitive to non-compliance with the homogeneity hypothesis, while ei_manual suffers more when the number of units decreases or when the complexity of the problem increases. It would be interesting to study if this insight could be used, when the convergence of the MCMC chains is guaranteed, to determine an optimal weight structure that maximizes the quality (accuracy) of an estimate based on a weighted mean.

Considering the previous idea further, and taking into account that, on the one hand, one of the main weaknesses of the approach implemented in ei.MD.bayes lies in the fact that its arguments need to be correctly tuned and, on the other hand, that nslphom usually produces reasonable solutions, although slightly worse than ei.MD.bayes solutions when this is properly tuned and converges, another line of research worth exploring would be to study whether the use of ei.MD.bayes could be automated by defining the priors of its Bayesian specification using the solution reached with nslphom. The outputs of nslphom could be employed to generate (overdispersed) priors for the ei.MD.bayes hyperparameters, including the possibility of using them to produce proper starting values for the α_{jk} and p^i_{jk}, which can be declared to ei.MD.bayes through its start.alphas and start.betas arguments.

The idea would be to study if this strategy would allow better solutions to be reached combining the strengths of both approaches in another way without paying the price of automation. Another advantage of this approach would be that it allows a more natural way of measuring the uncertainty of the estimates. Measures of uncertainty always relevant, that in some contexts, such as in US voting rights litigation, are extremely important. This approach, however, will not come without drawbacks. Using the nslphom output to define the ei.MD.bayes priors would not produce an authentic Bayesian estimate, since in this scenario the priors to be used by ei.MD.bayes would have been generated from the same data that it is going to employ to update them. In this case, this two-step strategy could be exclusively observed as an optimization method, but not as a proper Bayesian approach. Even though, using nslphom output to generate starting values for the MCMC chains does make sense, since it should lead to more efficient convergence and better tuning parameters.

In our discussion we have placed certain emphasis on automation (after all, we are dealing with a large number of elections) which is particularly relevant, for instance, in election night analysis. Nevertheless, depending on the context and the ultimate use of the estimates, making inferences beyond filling in the unobserved inner cells of the tables can be more than necessary (for example, in voting rights litigation or in academic studies), and this is more easily accomplished using a full statistical model than a mathematical programming algorithm. Because aggregation involves the loss of information at the individual level, any single approach to ecological inference requires some assumptions, with the success of the effort partially depending on these. Hence, in our view, it pays for the analyst to have a variety of methods that can be used depending on the purpose of the analysis and the logistic, human-resources and time constraints, and also for exploring the data. When different models lead to qualitatively similar conclusions, one can consider the results robust to the different sets of assumptions. But, when various models yield different conclusions, the analyst should, conditional on the ultimate aim of the estimates and/or the circumstances, examine the impact of the different assumptions on the conclusions or make her/his decisions with the aid of this and other comparative studies.

Acknowledgments

The authors wish to thank Carolina Plescia for providing us with the electoral Scottish data handled in this paper and two anonymous reviewers and the editors for their really valuable comments and suggestions. We are grateful to M. Hodkinson for translating and revising the English of the paper and Priscila Espinosa for her tips about $\LaTeX$. This research has been supported by Consellería d'Innovació, Universitats, Ciència i Societat Digital, Generalitat Valenciana [grant number AICO/2021/257] and by the Ministerio de Economía e Innovación [grant number PID2021-128228NB-I00].

Availability of data and material

The New Zealand data used in this research is publicly available on the website http://www.electionresults.org.nz. The Scottish data handled in this paper was provided by Carolina Plescia via personal communication. See also https://links.uv.es/72uQiop, DOI: 10.17605/OSF.IO/DY2SE.

Code availability

Using as a base some of the functions included in the R-packages `eiPack` (version 0.2-1) and `lphom` (version 0.1.3), the ad-hoc R-code employed to apply the assessed algorithms to the particular data analysed in this research is available, with comments in Spanish, in the html files available in https://links.uv.es/Htm570y, DOI: 10.17605/OSF.IO/ZAQH3.

References

Allport, F. H. (1924). The group fallacy in relation to social science. *American Journal of Sociology* 29(6), 688–706.

Barreto, M., Collingwood, L., Garcia-Rios, S., and Oskooii, K. A. R. (2022). Estimating candidate support in Voting Rights Act cases: Comparing iterative EI and EI-RC methods. *Sociological Methods & Research* 51, 271–304.

Brown, P. J., and Payne, C. D. (1986). Aggregate data, ecological regression, and voting transitions. *Journal of the American Statistical Association* 81(394), 452–460.

Choirat, C., Gandrud, C., Honaker, J., Imai, K., King, G., and Lau, O. (2017). *Zelig: Everyone's Statistical Software, Version 5.0-15*. URL: http://ZeligProject.org

Collingwood, L., Oskooii, K., Garcia-Rios, S., and Barreto, M. (2016). eiCompare: Comparing ecological inference estimates across EI and EI:RxC *The R Journal* 8, 92–101.

Dassonneville, R., and Hooghe, M. (2017). The noise of the vote recall question: The validity of the vote recall question in panel studies in Belgium, Germany, and the Netherlands. *International Journal of Public Opinion Research* 29(2), 316–338.

Ferree, K. E. (2004). Iterative approaches to RxC ecological inference problems: where they can go wrong and one quick fix. *Political Analysis* 12(2), 143–159.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review* 43, 557–572.

Forcina, A., and Pellegrino, D. (2019). Estimation of voter transitions and the ecological fallacy. *Quality & Quantity* 53, 1859–1874.

Gelman, A., Park, D.K., Ansolabehere, S., Price, L. C., and Minnite, L. C. (2001). Models, assumptions and model checking in ecological regression. *Journal of the Royal Statistical Society, Series A* 164(1), 101–118.

Gehlke, C. E., and Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association* 29(185A), 169–170.

Goodman, L. A. (1953). Ecological regressions and the behaviour of individuals. *American Sociological Review* 18, 663–664.

Goodman, L. A. (1959). Some alternatives to ecological correlation. *American Journal of Sociology* 64(6), 610–625.

Gosnell, H. F., and Gill, N. N. (1935). An Analysis of the 1932 Presidential vote in Chicago. *The American Political Science Review* 29, 967–984.

Greiner, D. J., and Quinn, K. M. (2009). RxC ecological inference: Bounds, correlations, flexibility, and transparency of assumptions. *Journal of the Royal Statistical Society, Series A* 172(1), 67–81.

Greiner, D. J., and Quinn, K. M. (2010). Exit polling and racial bloc voting: Combining individual-level and RxC ecological data. *The Annals of Applied Statistics* 4, 1774–1796.

Imai, K., King, G., and Lau, O. (2008). Toward a common framework for statistical analysis and development. *Journal of Computational and Graphical Statistics* 17, 892–913.

Johnston, R. J., and Hay, A. M. (1983). Voter transition probability estimates: An entropy-maximizing approach. *European Journal of Political Research* 11, 93–98.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

King, G., Rosen, O., and Tanner, M. A. (1999). Binomial-beta hierarchical models for ecological inference. *Sociological Methods & Research* 28, 61–90.

Klein, J. M. (2019). *Estimation of Voter Transitions in Multi-Party Systems. Quality of Credible Intervals in (hybrid) Multinomial-Dirichlet Models.* Master Thesis Dissertation. Ludwig-Maximilians-Universität München.

Klima, A., Thurner, P. W., Molnar, C., Schlesinger, T., and Küchenhoff, H. (2016). Estimation of voter transitions based on ecological inference: an empirical assessment of different approaches. *AStA - Advances in Statistical Analysis* 100, 133–159.

Klima, A., Schlesinger, T., Thurner, P. W., and Küchenhoff, H. (2019). Combining aggregate data and exit polls for the estimation of voter transitions. *Sociological Methods & Research* 48, 296–325.

Lau, O., Moore, O. R. T., and Kellermann, M. (2007). eiPack: RxC ecological inference and higher-dimension data management. *The R Journal* 7, 43–47.

Lau, O., Moore, O. R. T., and Kellermann, M. (2020). *eiPack: Ecological Inference and Higher-Dimension Data Management. R package version 0.2-1.* https://CRAN.R-project.org/package=eiPack

Manski, C. F. (2007). *Identification for Prediction and Decision*. Harvard University Press.

Martín, J. (2020). *Análisis de la incertidumbre en la estimación de la movilidad electoral mediante el procedimiento lphom*. PhD Dissertation. Universidad Politécnica de Valencia.

Miller, W. L. (1972). Measures of electoral change using aggregate data. *Journal of the Royal Statistical Society, Series A* 135, 122–142.

Ogburn, W. F., and Goltra, I. (1919). How women vote. *Political Science Quarterly* 34, 413–433.

Park, W., Hanmer, M. J., and Biggers, D. R. (2014). Ecological inference under unfavorable conditions: straight and split-ticket voting in diverse settings and small samples. *Electoral Studies* 36, 192–203.

Pavía, J. M. (2022). ei.Datasets: Real datasets for assessing ecological inference algorithms. *Social Science Computer Review* 40, 247–260.

Pavía, J. M. (2023). Adjustment of initial estimates of voter transition probabilities to guarantee consistency and completeness. *SN Social Sciences*, 3, 75.

Pavía, J. M., and Aybar, C. (2020). Electoral mobility in the 2019 elections in the Valencian region. *Debats. Journal on Culture, Power and Society*, 134, 27–51.

Pavía, J. M., and Romero, R. (2021). *lphom: Ecological Inference by Linear Programming under Homogeneity. R package version 0.1.3.* https://CRAN.R-project.org/package=lphom

Pavía, J. M., and Romero, R. (2022). Improving estimates accuracy of voter transitions. Two new algorithms for ecological inference based on linear programming. *Sociological Methods & Research*, online available, https://doi.org/10.1177%2F00491241221092725

Pawlowsky-Glahn, V., Egozcue, J. J., and Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data.* Chichester: John Wiley & Sons, Ltd.

Petropoulos, F., et al. (2022). Forecasting: Theory and practice. *International Journal of Forecasting* 38, 705–871.

Plescia, C., and De Sio, L. (2018). An evaluation of the performance and suitability of RxC methods for ecological inference with known true values. *Quality & Quantity* 52, 669–683.

Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer.

Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* 15, 351–357.

Romero, R., Pavía, J. M., Martín, J, and Romero, G. (2020). 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.

Romero, R., and Pavía, J. M. (2021). Estimating vote party entries and exits by ecological inference. Mathematical programming versus Bayesian statistics. *BEIO: Boletín de Estadística e Investigación Operativa* 34, 85–97.

Rosen, O., Jiang, W., King, G., and Tanner, M. A. (2001). Bayesian and frequentist inference for ecological inference: The RxC Case. *Statistica Neerlandica* 55(2), 134–156.

Roy, V. (2020). Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application* 7, 387–412.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society - Series B* 13(2), 238–241.

Thomsen, S. R. (1987). *Danish elections, 1920-79: A Logit Approach to Ecological Analysis and Inference.* Aarhus: Politica.

Upton, G. J. G. (1978). A note on the estimation of voter transition probabilities. *Journal of the Royal Statistical Society, Series A.* 141, 507–512.

Wakefield, J. (2004). Ecological inference for 2x2 tables (with discussion). *Journal of the Royal Statistical Society, Series A.* 167, 385–445.