



Adjustment of initial estimates of voter transition probabilities to guarantee consistency and completeness

Jose M. Pavía¹ 

Received: 24 March 2022 / Accepted: 2 April 2023 / Published online: 29 April 2023
© The Author(s) 2023

Abstract

The estimation of voter shifts (stayers and switchers) between elections is an active area of research that, for decades, has attracted the interest of many scholars. The voter transitions are typically summarised in a row-standardised proportion (probability) matrix. This matrix is usually unknown, despite it being of interest to many agents, including party teams, the media and political scientists. When surveys are used to approximate this matrix, it is not uncommon for the estimated matrix to be inconsistent and even incomplete. The iterative proportional fitting algorithm solves inconsistency but cannot fix incompleteness. Hierarchical Bayesian models that combine aggregate and survey estimates can solve both problems, but are extremely complex and data-demanding. This paper details all the scenarios concerning the available information that can be reasonably considered and, within the linear programming framework, develops specific models to reach consistency and completeness. The models are, moreover, quite flexible as they allow analysts to have missing values and to introduce through weights their relative confidences in the different a priori transition proportions. The usefulness of the proposed models is illustrated with real data. Interested readers can easily use these new models with their data as they have been programmed in the function `lp_apriori` of the R-package `lphom`.

Keywords Voter transitions · Linear programming · Elections · Surveys · Polls · `lphom`

Introduction

The estimation of voter shifts (stayers and switchers) between elections is an active area of research that has attracted the interest of many scholars for decades, both during the XX century (e.g., Vangrevelinghe 1961; Hawkes 1969; Irwin and Meeter 1969; McCarthy and Terence 1977; Upon 1978; Brown and Payne 1986; Tziafetas

✉ Jose M. Pavía
pavia@uv.es

¹ GIPEyOP, Universitat de Valencia, Valencia, Spain

1986; Thomsen 1987; Johnston and Pattie 1991 1993; Füle 1994) and the XXI century (e.g., Wellhofer 2001; Antweiler 2007; Park 2008; Andreadis and Chadjiapadelis 2009; Forcina et al. 2012; Park et al. 2014; Russo 2014; Puig and Ginebra 2014 2015; Klima et al. 2016 2019; Plescia and De Sio 2018; Klein 2019; Abou-Chadi and Stoetzer 2020; Pavía and Aybar 2020; Romero et al. 2020; Thurner et al. 2020; Pavía and Romero 2022a 2022c; Sandoval and Ojeda 2022; Thurner et al. 2022; Vizcaino and Pavía 2022).

The voter transitions among the J election options available in an election E1 held at time t and the K election options of another election E2, usually held at a posterior time $t + 1$, are typically summarised in a $J \times K$ row-standardised proportion (probability) matrix $\mathbf{P} = [p_{jk}]$; where p_{jk} represents the proportions of electors in the entire electoral space who chose (are classified in) option k in E2 among those who chose (are classified in) option j in E1. This matrix is usually unknown. Therefore, given their relevance to multiple agents (among others, party teams, the media and political scientists), it is routinely approximated using models and/or from polls (e.g., Klima et al. 2019; Abou-Chadi and Stoetzer 2020; Thurner et al. 2020).

When surveys are used to approximate this matrix, it is not uncommon for the estimated matrix to be inconsistent and even incomplete (e.g., Park et al. 2014; Russo 2014; Abou-Chadi and Stoetzer 2020). Inconsistency means that discrepancies emerge between actual results recorded in E2 and outcomes attained after applying the estimated probabilities to the results registered in E1. This is evidenced by examining the differences between estimated and real percentages. Incompleteness is a consequence of the unavailability of estimates for some proportions, either as a consequence of small sample sizes or because they are impossible to derive even from surveys. When this happens, analysts are usually interested in correcting these flaws, achieving a new estimated matrix: consistent and complete.

Two routes have been mainly followed to solve this problem: adjusting initial transfer probabilities using the Iterative Proportional Fitting (IPF) algorithm (e.g., Park 2008; Pavía and Aybar 2020; Thurner et al. 2020) or combining aggregate and a priori (survey) estimates within a statistical hierarchical Bayesian model (Greiner and Quinn 2010; Klima et al. 2019; Thurner et al. 2022). The first approach is quite simple, but it can only solve the inconsistency problem and is not free of weaknesses. The main limitation of IPF is its incapacity to move initial zero estimates, an issue that sometimes leads the algorithm to never converge (Thurner et al. 2020). In contrast, the second approach can solve both problems (inconsistency and incompleteness), but it is significantly more complex and data-demanding. Furthermore, when the data are available, Bayesian skill-trained analysts are still needed to properly tune the models.

This paper develops, within a linear programming framework, a family of models to reach in a fairly simple way consistency and completeness. All that is required is a (maybe incomplete) transfer vote matrix of initial estimates and two vectors of (row and column) constraints; these vectors usually correspond to the actual results recorded in E1 and E2 elections. The new models, which solve the incompleteness problem and are as simple to use as IPF, cover all the scenarios concerning the available a priori information (in terms of a priori proportions and their confidence) that can be reasonably considered. Interested analysts can easily use these new models

with the help of the function `lp_apriori` of the R-package `lphom` (Pavía and Romero 2022b).

The data

Without loss of generality, we assume that the aggregated electoral outcomes of a partition into I territorial units of the electoral space are known and that J and K are the total number of possible options/outcomes/situations in E1 and E2, respectively. In both cases, we consider abstention as a possible voting option as well as other situations, such as new entries or exits in the census. Although, in general, having the aggregate data of the entire electoral space (i.e., dealing with the case $I = 1$) is enough to solve the problem (see the examples in the supplementary material), having electoral outcomes of more units can be useful in the process of approximating (net) figures for entries and exits, when these are not available.

To solve the problem, we need to handle two sources of data. Aggregate recorded/estimated figures in each unit in both elections, $\mathbf{X} = [x_{ij}]$ and $\mathbf{Y} = [y_{ik}]$, and a priori information, $\mathbf{P}^0 = [p_{jk}^0]$.

On the one hand, for each of the $i = 1, \dots, I$ territorial units we need the (available) recorded/estimated statistics of voters/electors x_{ij} (for $j = 1, \dots, J_x$, with $J_x \leq J$) corresponding to the election options in E1 and, equally, we also need similar (available) figures y_{ik} (for $k = 1, \dots, K_y; K_y \leq K$) corresponding to E2. Note $J_x \leq J$ and/or $K_y \leq K$ because, sometimes, we will not have data on some collectives, such as census ins and outs.

On the other hand, we need a matrix $\mathbf{P}^0 = [p_{jk}^0]$ of order $J_0 \times K_0$ (where $J_0 \leq J$ and $K_0 \leq K$) of a priori transfers between elections 1 and 2. This matrix can come from a poll or from another source. In this matrix, p_{jk}^0 transfers are usually expressed in the form of row-standardised proportions, although a cross-classified matrix of counts, probably derived from a poll, is also allowed in our specifications. This matrix can contain missing values.

The basic unknowns are the quantities p_{jk} . The goal is to derive a consistent and complete row-standardised matrix of proportions, $\hat{\mathbf{P}}$, as close as possible to \mathbf{P}^0 . To do this, we consider eight different scenarios (some of them with several variants) relevant to the information available in \mathbf{X} and \mathbf{Y} . To the five scenarios defined in Romero et al. (2020)—simultaneous, raw, regular, full and gold—we add three additional scenarios: ordinary, enriched and semifull.

As in Romero et al. (2020), we consider that entries and exits of the census can each have two different sources. On the one hand, entries in each territorial unit i (E_i) are the sum of two groups: young electors newly entitled to vote (N_i) and new residents (immigrants, I_i) that have the right to vote. On the other hand, exits (X_i) are also made up of two groups: voters registered in E1 who have died before E2 (D_i) and people who have emigrated during the inter-election period (M_i). The scenarios (and their variants) basically differ regarding the information available for these groups. In raw, regular, ordinary and enriched scenarios, the row-aggregations of

X and Y will typically differ and data about net entries and/or net exits should be derived from the available information to guarantee congruence.

In general, denoting the total censuses corresponding to unit i in both elections by $C1_i$ and $C2_i$, the following accounting equalities can be expressed:

$$C1_i + E_i - X_i = C2_i \iff C1_i + (N_i + I_i) - (D_i + M_i) = C2_i$$

Hence, given that $C1_i$ and $C2_i$ are known, depending on the level of information available regarding the remaining components, we can estimate: $b_i = E_i - X_i$, whose absolute value corresponds to either total net entries, if $b_i > 0$, or to total net exits, if $b_i < 0$; $c_i = N_i - D_i$, whose modulus corresponds to either net entries different from immigrants, if $c_i > 0$, or net exits different from emigrants, if $c_i < 0$; or $d_i = I_i - M_i$, which without sign corresponds to either net entries different from new voters by age, if $d_i > 0$, or net exits different from deaths, if $d_i < 0$. The aggregation by units of these quantities allows estimates of the sizes of these collectives in the entire population to be achieved. In general, we recommend working with small units to avoid (in absolute values) underestimations of b_i 's and c_i 's and with large units to avoid overestimations of d_i 's. In any case, since, as a rule, (i) these groups tend to be marginal, (ii) the focus is not usually on their electoral behaviour and (iii) it is usually simple to obtain accurate estimates of Y_i and M_i from demographic statistics (see, e.g., Pavía and Veres-Ferrer 2016a, b), it is advisable to work with large units. This also has the advantage of reducing the costs of data wrangling (Klima et al. 2016).

The scenarios

In this section, we describe the characteristics of the different scenarios and detail how the data should be arranged in each one of them. In all the cases, we assume that the rows of P^0 and the columns of X are sorted in the same order for $\min(J_0, J_x)$ and that this also happens between the columns of P^0 and columns of Y for $\min(K_0, K_y)$. These arrangements are transferred to \hat{P} .

So as keep the complex discussion of scenarios simple and in order not to include P^0 in the discussion, we assume without loss of generality that (i) the matrix P^0 has been derived from a poll (i.e., no information is enclosed in P^0 about exits and therefore $K_0 \leq K_y$) and (ii) minor voting options are grouped in X , Y and P^0 in an others option. Although no hypothesis is stated regarding the relationship between J_0 and J_x , in general $J_0 \geq J_x$ when information about the electoral behaviour of new voters is available.

Simultaneous scenario

This is the simplest scenario. In this case the same electors are entitled to vote in E1 and E2, usually because both elections are held at the same time (Pavía and Romero 2022c). In this scenario the sum by rows X and Y must coincide and special constraints about the coefficients of the matrix (such as the ones stated in Eqs. (6)–(15) introduced in Sect. 5) do not apply. Here $J = J_x$ and $K = K_y$ and the figures N_i, I_i, D_i

and M_i are null by definition, so $b_i = c_i = d_i = 0$. This defines a basic scenario (type I), which can be solved with the basic model defined by Eqs. (1)–(5); a model with no specific proportions' constraints.

Raw scenario

This scenario accounts for the most common situation. A situation with two elections elapsed over some time and where only the raw election data recorded in the I territorial units in which the area under study is divided are available. In this scenario, net exits (deaths and emigrants) and net entries (new young voters and immigrants) are estimated using the b_i 's coefficients. We can have four variants, depending on the values of the b_i 's. If $b_i = 0 \forall i$, this case collapses to a simultaneous scenario. If $b_i \geq 0 \forall i$, with at least a b_i strictly positive, $J = J_x + 1$ and $K = K_y$ and the resulting problem can be also solved using the basic model. If $b_i \leq 0 \forall i$, with at least a b_i strictly negative, $J = J_x$ and $K = K_y + 1$. In this case, it seems reasonable to follow Romero et al. (2020) and to assume that exits (mainly as a consequence of mortality) impact uniformly in all the political options: $p_{1,K} = p_{2,K} = \dots = p_{J,K}$. We call this kind of scenario type II. If the uniform hypothesis is not assumed, the scenario collapses to a type I scenario. Finally, if strictly positive and negative b_i 's are obtained, then $J = J_x + 1$ and $K = K_y + 1$. In this case, the J^{th} option of E1 will correspond to (net) new entries and the K^{th} option of E2 to (net) exits. Since new entries cannot be exits, the logical constraint $p_{J,K} = 0$ applies. This gives rise to two new types of scenarios. If the uniform hypothesis is assumed ($p_{1,K} = p_{2,K} = \dots = p_{J-1,K}$), the type III scenario appears, otherwise this gives a type IV scenario.

Regular scenario

Initially, this scenario looks similar to the raw scenarios; it only differs in the information provided regarding E1, which is enlarged. In this case, the data contained in the last column of X refers to, without loss of generality, new young electors who have the right to vote for the first time (formally there is no differentiation in the discussion as to whether this refers to immigrants or new young electors). Here, again, new variants arise depending on the information we can derive for net entries (different from new young electors) and net exits. Without net entries and net exits, we are formally in a type I scenario, with $J = J_x$ and $K = K_y$. With net exits but not net entries, we have $J = J_x$ and $K = K_y + 1$ and we are in either a type III scenario if the uniform hypothesis is assumed or in a type IV scenario if that hypothesis is not assumed. With net entries but not net exits, $J = J_x + 1$ and $K = K_y$ and we again are in a type I scenario. Finally, if we have both net entries and exits, we need to impose by logic the constraints $p_{J-1,K} = p_{J,K} = 0$. In this last case, where $J = J_x + 1$ and $K = K_y + 1$, we can consider two variants depending on whether or not the uniform hypothesis (that this time corresponds to $p_{1,K} = p_{2,K} = \dots = p_{J-2,K}$) is imposed.

The last case defines two new basic types of scenarios. If the uniform hypothesis is imposed, we are in a type V scenario and if it is not imposed, in a type VI scenario.

Ordinary scenario

This scenario is also an extension of a raw scenario but this time with more information available regarding E2. In this case, the data contained in the last column of Y refer to, without loss of generality, exits of the census due to death (formally there is no differentiation in the discussion if it refers to emigrants). As usual, new variants arise depending on the information we can derive for net entries and net exits (different from deaths). Without net entries and net exits, $J = J_x$ and $K = K_y$ and we are formally in either a type II scenario if the uniform hypothesis about exits is imposed or in a type I scenario without this hypothesis. With net entries but not net exits, $J = J_x + 1$ and $K = K_y$ and we are in either a type III scenario with the uniform hypothesis or in a type IV without it. With net exits but not net entries, we have $J = J_x$ and $K = K_y + 1$ and we are in either a type I scenario if the uniform hypothesis is not assumed or in a new type scenario if it is assumed; we call this scenario type VII. In type VII scenarios, we have two possible types of exits (death and emigration) and the uniform hypothesis applies for both of them: $p_{1,K-1} = p_{2,K-1} = \dots = p_{J,K-1}$ and $p_{1,K} = p_{2,K} = \dots = p_{J,K}$. Finally, if we have both net entries and exits, the constraints $p_{J,K-1} = p_{J,K} = 0$ need to be imposed, $J = J_x + 1$, $K = K_y + 1$, and two new basic types of scenarios, depending on whether or not the uniform hypothesis (that this time crystallizes into $p_{1,K-1} = p_{2,K-1} = \dots = p_{J-1,K-1}$ and $p_{1,K} = p_{2,K} = \dots = p_{J-1,K}$) is assumed. If this hypothesis is assumed, we are in a type VIII scenario and if it is not assumed, in a type IX scenario.

Enriched scenario

This scenario extends raw scenarios in the two directions analysed in regular and ordinary scenarios. Here, we have without loss of generality data of new young electors by age in the last column of X and of deaths in the last column of Y . Again, different variants arise depending on whether net additional entries (due to immigration) and/or net additional exits (due to emigration) are estimated. Without net entries nor net exits, we are in a type III scenario if uniform hypothesis is assumed and in a type IV scenario without it. With net entries but not net exits, the scenarios that emerge are type V and type VI with and without the uniform hypothesis, respectively. With net exits but not net entries, the types of scenarios that arise are of types VIII and IX with and without the uniform hypothesis, respectively. Finally, with both net entries and net exits, two new basic scenarios emerge. In this case, it is necessary to include by logic $p_{J-1,K-1} = p_{J-1,K} = p_{J,K-1} = p_{J,K} = 0$ and also $p_{1,K-1} = p_{2,K-1} = \dots = p_{J-2,K-1}$ and $p_{1,K} = p_{2,K} = \dots = p_{J-2,K}$ if the uniform hypothesis is assumed. We call the scenario with all the above constraints type X, and type XI the corresponding scenario only containing the logic constraints.

Semifull scenario

In this scenario, the analyst has both aggregate information about total entries (young electors and new immigrants) and total exits (deaths and emigrants) of the census list. Total entries and exits are assumed to be in the last columns of X and Y , respectively. Here, the sum by rows of X and Y must agree and there are only two variants. If the uniform hypothesis is assumed we are in a type III scenario and in a type IV, otherwise.

Full scenario

In this scenario, the analyst has detailed information about totals of new young electors that have the right to vote for the first time (penultimate column of X), and of new immigrants that have the right to vote (last column of X) and aggregate information about total exits (due to death or emigration) of the census lists (last column of Y). Here, again, the sum by rows of X and Y must agree and there are only two variants. If the uniform hypothesis is assumed we are in a type V scenario and, if not, in a type VI.

Gold scenario

This scenario is similar to a full scenario but here total exits are separated out between exits due to emigration and due to death (penultimate and ultimate columns of Y). Again, the sum by rows of X and Y must agree and there are only two variants, depending on whether or not the uniform hypothesis is assumed. We are in a type X scenario under this hypothesis and in a type XI otherwise.

Schematic representation of scenarios

The above discussion is quite dense. We have considered eight different scenarios regarding the information available for X and Y and, after considering different variants, this gives us thirty-five possibilities that collapse in eleven basic structures for the estimated transfer matrix. Table 1 schematically summarises the cases discussed.

The model

Whatever the scenario and linked constraints, our model reaches its solutions after solving a linear programming model whose objective is to minimize a weighted sum of the absolute value discrepancies between the pairs p_{jk} and p_{jk}^0 . The model imposes the consistency property (Eqs. (1)–(3)) to the solution and does not necessarily deal with all the discrepancies, defined in Eq. (4), symmetrically. They are weighted in the objective function (see Eq. (5)). Because the level of confidence for all the a priori proportions, p_{jk}^0 , is not usually the same, we inform the model about this using

Table 1 Summary of the scenarios included in the *lp_apriori* function

Scenarios	Associated original scenarios, after variants	Constraints	Matrix structure
Type I	Simultaneous Raw, no entries, no exits Raw, entries, no exits Raw, no entries, exits, no uniform Regular, no entries, no exits Regular, entries, no exits Ordinary, no entries, no exits, no uniform Ordinary, no entries, exits, no uniform	–	$\begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}$
Type II	Raw, exits, no entries, uniform Ordinary, no entries, no exits, uniform	$p_{1,K} = p_{2,K} = \dots = p_{J,K}$	$\begin{bmatrix} ? & ? & ? & * \\ ? & ? & ? & * \\ ? & ? & ? & * \\ ? & ? & ? & * \end{bmatrix}$
Type III	Raw, exits, entries, uniform Regular, no entries, exits, uniform Ordinary, entries, no exits, uniform Enriched, no entries, no exits, uniform Semifull, uniform	$p_{1,K} = p_{2,K} = \dots = p_{J-1,K}$ $p_{J,K} = 0$	$\begin{bmatrix} ? & ? & ? & * \\ ? & ? & ? & * \\ ? & ? & ? & * \\ ? & ? & ? & 0 \end{bmatrix}$
Type IV	Raw, entries, exits, no uniform Regular, no entries, exits, no uniform Ordinary, entries, no exits, no uniform Enriched, no entries, no exits, no uniform Semifull, no uniform	$p_{J,K} = 0$	$\begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & 0 \end{bmatrix}$
Type V	Regular, entries, exits, uniform Enriched, entries, no exits, uniform Full, uniform	$p_{1,K} = p_{2,K} = \dots = p_{J-2,K}$ $p_{J-1,K} = p_{J,K} = 0$	$\begin{bmatrix} ? & ? & ? & * \\ ? & ? & ? & * \\ ? & ? & ? & * \\ ? & ? & ? & 0 \\ ? & ? & ? & 0 \end{bmatrix}$
Type VI	Regular, entries, exits, no uniform Enriched, entries, no exits, no uniform Full, no uniform	$p_{J-1,K} = p_{J,K} = 0$	$\begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & 0 \\ ? & ? & ? & 0 \end{bmatrix}$
Type VII	Ordinary, no entries, exits, uniform	$p_{1,K-1} = p_{2,K-1} = \dots = p_{J,K-1}$ $p_{1,K} = p_{2,K} = \dots = p_{J,K}$	$\begin{bmatrix} ? & ? & - & * \\ ? & ? & - & * \\ ? & ? & - & * \\ ? & ? & - & * \end{bmatrix}$
Type VIII	Ordinary, entries, exits, uniform Enriched, no entries, exits, uniform	$p_{1,K-1} = p_{2,K-1} = \dots = p_{J-1,K-1}$ $p_{1,K} = p_{2,K} = \dots = p_{J-1,K}$ $p_{J,K-1} = p_{J,K} = 0$	$\begin{bmatrix} ? & ? & - & * \\ ? & ? & - & * \\ ? & ? & - & * \\ ? & ? & 0 & 0 \end{bmatrix}$
Type IX	Ordinary, entries, exits, no uniform Enriched, no entries, exits, no uniform	$p_{J,K-1} = p_{J,K} = 0$	$\begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & 0 & 0 \end{bmatrix}$

Table 1 (continued)

Scenarios	Associated original scenarios, after variants	Constraints	Matrix structure
Type X	Enriched, entries, exits, uniform Gold, uniform	$p_{1,K-1} = p_{2,K-1} = \dots = p_{J-2,K-1}$ $p_{1,K} = p_{2,K} = \dots = p_{J-2,K}$ $p_{J-1,K-1} = p_{J-1,K} = 0$ $p_{J,K-1} = p_{J,K} = 0$	$\begin{bmatrix} ? & ? & - & * \\ ? & ? & - & * \\ ? & ? & - & * \\ ? & ? & 0 & 0 \\ ? & ? & 0 & 0 \end{bmatrix}$
Type XI	Enriched, entries, exits, no uniform Gold, no uniform	$p_{J-1,K-1} = p_{J-1,K} = 0$ $p_{J,K-1} = p_{J,K} = 0$	$\begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & 0 & 0 \\ ? & ? & 0 & 0 \end{bmatrix}$

Source: Own elaboration

weights, $\omega_{jk} > 0$. In particular, the basic model is defined by the system of Eqs. (1)–(5).

$$p_{jk} \geq 0 \text{ for } j = 1, \dots, J \text{ } k = 1, \dots, K \tag{1}$$

$$\sum_{k=1}^K p_{jk} = 1 \text{ for } j = 1, \dots, J \tag{2}$$

$$\sum_{j=1}^J p_{jk} x_j = y_{\cdot k} \text{ for } k = 1, \dots, K \tag{3}$$

$$(p_{jk}^0 - p_{jk}) = e_{jk}^+ - e_{jk}^- \text{ for } j = 1, \dots, J \text{ } k = 1, \dots, K \tag{4}$$

$$\text{minimize } Z = \sum_{j=1}^J \sum_{k=1}^K \omega_{jk} (e_{jk}^+ + e_{jk}^-) \tag{5}$$

where $x_j = \sum_{i=1}^I x_{ij}$ and $y_{\cdot k} = \sum_{i=1}^I y_{ik}$.

Although for simplicity in the above mathematical representation of the problem it has been stated that Eqs. (4) and (5) range for all the values of j and k , in practice they are just delimited for the pair of indexes (j, k) for which p_{jk}^0 and ω_{jk} are defined.

In addition to the above equations, depending on the type of scenario we are in, the system will also include (except in type I scenarios) some of constraints defined in Eqs. (6)–(15). The interested reader can find in Table 1 the relationships between the scenarios and their associated additional constraints.

$$p_{1,K} = p_{2,K} = \dots = p_{J,K} \tag{6}$$

$$p_{1,K-1} = p_{2,K-1} = \dots = p_{J,K-1} \quad (7)$$

$$p_{1,K} = p_{2,K} = \dots = p_{J-1,K} \quad (8)$$

$$p_{1,K-1} = p_{2,K-1} = \dots = p_{J-1,K-1} \quad (9)$$

$$p_{1,K-1} = p_{2,K-1} = \dots = p_{J-2,K-1} \quad (10)$$

$$p_{1,K} = p_{2,K} = \dots = p_{J-2,K} \quad (11)$$

$$p_{J-1,K-1} = 0 \quad (12)$$

$$p_{J-1,K} = 0 \quad (13)$$

$$p_{J,K-1} = 0 \quad (14)$$

$$p_{J,K} = 0 \quad (15)$$

The weights

The weight $\omega_{jk} > 0$ is the penalty that we assign to the discrepancy $|p_{jk}^0 - p_{jk}|$, either e_{jk}^+ or e_{jk}^- , in Eq. (5). This coefficient measures the (relative) degree of confidence we have in the a priori value p_{jk}^0 . Everything else being constant, the greater this weight the closer the estimated p_{jk} and the p_{jk}^0 proportions will be. This means that setting the weight ω_{jk} to an arbitrarily large number will be, in many practical situations, equivalent to including in the model an additional restriction, the constraint $p_{jk} = p_{jk}^0$.

Although the weights can be exogenously stated by the user and introduced in the model through a matrix of weights, $\mathbf{W} = [\omega_{jk}]$, the `lp_apriori` function also has programmed functionalities derived from the contextual information. In addition to the possibility of introducing personal weights through a matrix, `lp_apriori` includes seven more ways of computing the weights, denoted by "constant", "x", "xy", "expected", "counts", "sqrt" and "sd".

When weights are set to "constant", the same credibility is attached to all the a priori proportions and, by default, all the weights are set equal to 1. In practice, however, analysts tend to have more information from proportions related to the largest groups. For example, a survey is more likely to interview people who belong to the groups whose election options are more frequent. The "x", "xy" and "expected" strategies to compute the weights exploit this fact by implementing different

approaches. In particular, the "x", "xy" and "expected" weights are calculated as proportional to, respectively, $\omega_{jk} \propto x_{.j}$, $\omega_{jk} \propto x_{.j}y_{.k}$ and $\omega_{jk} \propto x_{.j}p_{jk}^0$.

The a priori information is typically introduced to `lp_apriori` through a row-standardised matrix of proportions, but it could also be informed via a matrix, $[n_{jk}]$, of counts. When this happens, `lp_apriori` internally transforms the a priori data into a row-standardised matrix of proportions, but the issue here is that we can also use this information to define the weights. In this case, weights are defined observing the (underlying) sampling properties of the counts. The "counts" and "sqrt" weights are computed respectively proportional to $n_{jk} + \frac{1}{2}$ and $\sqrt{n_{jk} + 0.5}$, with the definition of "sd" weights being slightly more complex. They are $\omega_{jk} \propto \sqrt{\frac{n_{jk} + 0.5}{p_{jk}^0(1-p_{jk}^0)}}$ if $p_{jk}^0 \neq 0$ and $\omega_{jk} \propto \sqrt{2}$ if $n_{jk} = p_{jk}^0 = 0$.

Assessing the models with real data

This section assesses, using real data, whether the adjustment corrects the biases introduced by polls when estimating vote transfer matrices and to what extent. The impact on the bias reduction of the sample sizes and the different weight structures considered in the previous section are also analysed. To do this, a realistic random sample of each of the 565 elections available in the R-package `ei.Datasets` (Pavía 2022) is simulated and their corresponding sample-estimated and model-adjusted transition matrices compared to the actual ones.

The main difficulty when assessing methods to estimate vote transfer matrices between elections lies in the fact that, due to the secret ballot, actual cross-distributions of votes are as a rule unknown. In simultaneous elections, however, when the same electors cast their votes in the same ballot for several elections, actual vote transfer matrices can be collected. The `ei.Datasets` database gathers the 565 real crosstabs of votes corresponding to the party-to-candidate cross-distributions recorded in the 492 electorates of the New Zealand general elections held between 2002 and 2020 and in the 73 constituencies of the 2007 Scottish Parliament election. Hence, the `ei.Datasets` crosstabs are exploited to answer the previous questions.

A random sample of size n (with $n = 250, 500$ or 1000) is simulated from each election, assuming that polls suffer from both non-response bias and response error. Given that the average size of the populations in `ei.Datasets` is $\sim 33,192$ voters, sample sizes of this order seem reasonable. Indeed, the minimum sample size recommended in NZ for electorate/regional polls is 250 (Research Association New Zealand 2020), 500 being the standard sample size. Likewise, it is reasonable to assume that polls are going to be impacted both by differential response rates (non-response bias), which depend on voters' preferences and characteristics, and nonresponse error, due to social desirability issues or inaccurate recall voting. Actually, there is a large stream of literature documenting this impact occurring in real-world surveys all around the globe (see, e.g., Groves et al. 2002; Pavía et al. 2016; Cavari and Freedman 2022). The simulated samples are summarised in two-way contingency

tables with parties in rows and candidates in columns and small parties and candidates (those who gain less than a 3% of the votes in the election) grouped under the “others” option. The row-standardised versions of these tables correspond to the sample-estimated transition matrices.

The estimated transition matrices are now adjusted to match the actual observed distributions of votes recorded in each election. The adjustments are made with the help of the `lp_apriori` function of the `lphom` package, testing as weights the seven precomputed possibilities available to them in the function: “constant”, “x”, “xy”, “expected”, “counts”, “sqrt” and “sd”. At the end of this process, we have eight estimated vote transition matrices for each election: the one that comes directly from the poll and the seven attained after adjusting this. These matrices are compared with actual transitions using three distance statistics: *EI*, *EPW* and *MAN*, given by Eqs. (16)–(18). The interested reader can replicate all the computations and find more in-depth information by reviewing the R reproducible code available in the supplementary material (see the Data availability statement).

$$EI = 100 \cdot \frac{0.5 \sum_{j=1}^J \sum_{k=1}^K |v_{jk} - \hat{v}_{jk}|}{\sum_{j=1}^J \sum_{k=1}^K v_{jk}} \quad (16)$$

$$EPW = 100 \cdot \frac{\sum_{j=1}^J \sum_{k=1}^K v_{jk} |p_{jk} - \hat{p}_{jk}|}{\sum_{j=1}^J \sum_{k=1}^K v_{jk}} \quad (17)$$

$$MAN = 100 \cdot \frac{\sum_{j=1}^J \sum_{k=1}^K |p_{jk} - \hat{p}_{jk}|}{JK} \quad (18)$$

where v_{jk} and \hat{v}_{jk} stand for the actual and estimated/adjusted number of voters, respectively, voting simultaneously for party j and candidate k ; and p_{jk} and \hat{p}_{jk} represent the actual and estimated/adjusted proportions of voters who, respectively, vote for candidate k among those who vote for party j .

While *EI* measures distances between transfer matrices of votes, *EPW* and *MAN* deal with row-standardised transition matrices. *EI* accounts for the percentage of wrongly assigned votes in the estimated matrix: the minimum percentage of votes that should be moved among the cells of the table to reach a perfect fit. *MAN* measures the mean of the differences between the actual and the estimated proportion transitions, and *EPW* a similar distance to *MAN*, but with the individual differences weighted by the number of votes corresponding to the transfer between party j and candidate k . The smaller the numbers of these statistics, the closer the estimated/adjusted and the actual matrices.

Before presenting and analysing the results of the full simulation exercise, the whole process described above is first illustrated through an example. Consider the electorate of Maungakiekie in the 2020 NZ general election, the actual outcomes of which are presented in Table 2. Note that looking at the second panel of Table 2

Table 2 Actual Outcomes recorded in Maungakiekie during the 2020 NZ general election

Parties	Candidates					
	Fer. (ACT)	Men.(GP)	Lee (NP)	Rad. (LP)	Others	
Actual transfer matrix of votes						
ACT New Zealand (ACT)	532	17	1974	88	39	2650
Green Party (GP)	10	1084	397	1871	20	3382
Labour Party (LP)	220	1230	3395	13,325	560	18,730
National Party (NP)	347	59	8836	155	224	9621
Others parties	116	276	995	793	694	2874
	1225	2666	15,597	16,232	1537	
Actual row-standardised vote transition matrix						
ACT New Zealand (ACT)	20.08	0.64	74.49	3.32	1.47	7.11
Green Party (GP)	0.30	32.05	11.74	55.32	0.59	9.08
Labour Party (LP)	1.17	6.57	18.13	71.14	2.99	50.27
National Party (NP)	3.61	0.61	91.84	1.61	2.33	25.82
Others parties	4.04	9.60	34.62	27.59	24.15	7.71
	3.29	7.16	41.86	43.57	4.13	

The central cells of the upper and lower panels of the table contain the actual transfer matrix of votes and the corresponding row-standardised transition matrix recorded in the election, respectively. The actual distributions of votes gained by the parties in votes and percentages are displayed in the last column of the upper and lower panels, respectively. The equivalent distributions for candidates are presented in the last rows of each panel

Table 3 Simulated sample of size 500 in Maungakiekie for the 2020 NZ general election

Parties	Candidates					
	Fer. (ACT)	Men.(GP)	Lee (NP)	Rad. (LP)	Others	
Transfer matrix of votes of a simulated sample of size 500						
ACT New Zealand (ACT)	8	3	28	2	2	8.60
Green Party (GP)	5	23	10	14	0	10.40
Labour Party (LP)	6	20	54	95	5	36.00
National Party (NP)	20	4	136	1	6	33.40
Others parties	5	10	22	8	13	11.60
	8.80	12.00	50.00	24.00	5.20	
Row-standardised vote transition matrix linked to the sample of the upper panel						
ACT New Zealand (ACT)	18.60	6.98	65.12	4.65	4.65	6.98
Green Party (GP)	9.62	44.23	19.23	26.92	0.00	9.87
Labour Party (LP)	3.33	11.11	30.00	52.78	2.78	47.16
National Party (NP)	11.98	2.40	81.44	0.60	3.59	26.06
Others parties	8.62	17.24	37.93	13.79	22.41	10.22
	7.63	12.05	45.41	30.53	4.38	

The central cells of the upper and lower panels of the table contain, respectively, the transfer matrix of votes and the corresponding row-standardised transition matrix obtained after simulated a realistic poll of size 500. The raw sample party and candidate distributions in percentage are displayed, respectively, in last column and last row of the upper and lower panels, while the corresponding transfer matrix distributions corresponding are presented in the last column and row of the lower panel

there is evidence of strategic voting in this election and of a large number of switchers; an issue that emphasizes the importance of knowing the transition matrix (Abou-Chadi and Stoetzer 2020). For instance, 55.32% of the electors who voted for the Green Party also chose to vote in the ballot for the Labour Party candidate (Priyanca Radhakrishnan) and 74.49% of ACT voters chose to vote for the National Party candidate (Denise Lee).

Let us consider a random sample of size 500 selected from this electorate, as summarised in Table 3. By comparing the distributions in the last column and in the last row of the upper panel of Table 3 with the equivalent distributions in the lower panels of Table 2, it can be seen that, although the sample captures the general trends of the election, its raw estimates miss the actual results by a large amount. This sample shows a relevant level of bias.

Furthermore, although the estimates of the marginal distributions can be improved using post-stratification, that is, when the sample party-to candidate (candidate-to-party) transition matrix is used to approximate the candidate (party) distributions conditional on the actual party (candidate) distributions (as can be done in non-simultaneous elections), significant bias still remains in the marginal estimates. This can be confirmed comparing the distributions in the last column and the last row of the lower panel of Table 3 with the equivalent distributions in the lower panels of Table 2.

If the focus is on the transition matrix, as in our case, we see that its estimation can be improved by adjusting it, i.e., by making it consistent with the actual outcomes. This is evident since the transition matrix in Table 4, obtained after adjusting the data in Table 3 using $lp_apriori$ with $weights = "x"$, is closer to the actual transition matrix (displayed in the lower panel of Table 2) than the equivalent sampling transition matrix (see the lower panel of Table 3). For instance, in this example the *MAN* distance is reduced from 6.83 to 5.08.

The above example, which corresponds to one of the scenarios that can be replicated using the code of the supplementary material (see the Data availability statement), is quite representative of the full set of scenarios, since 90.44%, 93.81% and

Table 4 Adjusted matrix with "x" weights corresponding to the sample in Table 3

Transfer matrix of votes of a simulated sample of size 500

Parties	Candidates					
	Fer. (ACT)	Men. (GP)	Lee (NP)	Rad. (LP)	Others	
ACT New Zealand (ACT)	18.60	6.98	65.13	4.64	4.64	7.11
Green Party (GP)	0.00	44.23	19.22	36.35	0.00	9.08
Labour Party (LP)	0.00	1.38	22.94	73.42	2.26	50.27
National Party (NP)	5.03	2.39	81.44	7.55	3.60	25.82
Others parties	8.63	17.26	37.93	13.78	22.41	7.71
	3.29	7.16	41.86	43.57	4.13	

The percentage party and candidate distribution corresponding to the adjusted transfer matrix of votes are displayed in last column and row of the table, respectively

Table 5 Average of EI, EPW and MAN distances for the simulated samples by sample size

Weights	Sample size: 250			Sample size: 500			Sample size: 1000		
	EI	EPW	MAN	EI	EPW	MAN	EI	EPW	MAN
constant	9.75	5.49	6.80	8.71	4.98	5.87	7.99	4.61	5.35
x	9.66	5.38	6.75	8.59	4.87	5.78	7.84	4.50	5.24
xy	11.43	6.90	7.75	10.06	6.08	6.77	9.43	5.70	6.29
expected	14.79	8.94	9.04	14.08	8.48	8.39	13.47	8.08	7.98
counts	13.33	7.65	7.95	12.66	7.24	7.25	12.39	7.07	6.95
sqrt	10.77	5.65	7.02	9.81	5.06	6.13	9.12	4.63	5.65
sd	9.89	5.40	6.80	8.91	4.88	5.90	8.17	4.43	5.38
Sample	12.59	8.57	7.45	11.42	8.00	6.49	10.62	7.56	5.94

The smaller the numbers of these statistics, the closer the estimated/adjusted and the actual matrices

88.50% of the matrix adjustments made with weights="x" are closer (as measured with *EI*, *EPW* and *MAN*, respectively) to the actual matrix than the sample matrix. Indeed, as can be seen in Table 5, where the averages of the distance statistics recorded for the 565 elections is shown grouped by weights used in *lp_aprioi* and sample size, the adjustment of the raw sample matrix leads as a rule to more accurate matrix estimates.

On average, the results show that adjusting with weights="x" generates the most accurate solutions, whereas adjusting with weights="expected" worsens the sampling estimates. As weights="expected" gives more credibility to the sampling estimates with the largest expected number of votes, $v_{jk}^0 = x_{jk}p_{jk}^0$, this weighting exacerbates the sampling bias. Despite this result with weights="expected", adjusting leads to more accurate solutions on average for five out of the seven weightings ("constant", "x", "xy", "sqrt" and "sd"). The solutions achieved using weights "x" and "sd" are the most accurate. These results, moreover, are consistent by sample size: the same conclusions are reached regardless of the sample size. The impact of the sample size is seen in an improvement in the poll estimates, as expected, and also the adjustments. Indeed, as the sample size grows, both the sample and adjusted estimates improve in the same proportion.

Conclusions

This paper describes a family of models to adjust initial estimates of row-standardised voter transition probabilities to reach consistency and completeness in thirty-five different situations. All of them can be solved with the function *lp_apriori* available in the R-package *lphom* (Pavía and Romero 2022b). This package provides several algorithms based on linear programming for estimating, under the homogeneity hypothesis, general $J \times K$ ecological contingency tables and, in particular, vote transition matrices.

Although the `lp_apriori` function has been included in the `lphom` package, it should be noted that the solutions programmed in `lp_apriori` are conceptually different from the rest of the procedures available in the package. While the models in `lp_apriori` have been conceived to adjust (initial) estimates by modifying the available estimates as little as possible, making them consistent and complete, the rest of the algorithms of the package have been devised to generate estimates from aggregate data by employing the homogeneity hypothesis. They all, however, share the same mathematical approach to solving the problem (linear programming) and the aim of estimating a transfer matrix of votes.

The suggested models are not only valuable in their own right, as the previous section shows, but they also open the way to solving one of the limitations of the `lphom`-family algorithms. According to Greiner (2007, p. 120), “... a good [ecological inference] method should be flexible enough to incorporate information from surveys or exit polls” and while this issue has been recently addressed within the Bayesian ecological inference framework (Greiner and Quinn 2010; Klima et al. 2019) it still remains to be solved within the linear programming ecological inference approach; a matter which deserves prompt attention since its methods are much easier to apply. The new models introduced in this paper will make it possible to develop new methods capable of integrating a priori information and aggregate results within the linear programming framework.

Acknowledgements The authors wish to thank M. Hodkinson for revising the English of the paper. The author acknowledges the support of Generalitat Valenciana through project AICO/2021/257 (Conselleria d’Innovació, Universitats, Ciència i Societat Digital) and of Ministerio de Economía e Innovación through project PID2021-128228NB-I00.

Author contributions The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Generalitat Valenciana (Conselleria d’Innovació, Universitats, Ciència i Societat Digital), grant number AICO/2021/257, and Ministerio de Economía e Innovación, grant number PID2021-128228NB-I00 funded the research.

Data availability The data used to assess the models is publicly available on the R-package `ei.Datasets` (version 0.0.1–1) accessible on CRAN in the URL <<https://CRAN.R-project.org/package=ei.Datasets>> and the supplementary material is publicly available in the URL <<https://osf.io/gn83q/>>, <https://doi.org/10.17605/OSF.IO/GN83Q>. The supplementary material includes the reproducible ad hoc R-code employed, based on functions available in the R-package `lphom` (version 0.3.0–7), as well as the outputs (datasets) necessary to replicate and build upon the findings reported in the article.

Declarations

Conflict of interest The author states that there is no conflict of interest.

Ethical approval Ethics approval was not required for this study.

Informed consent This article does not contain any studies with human participants performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abou-Chadi T, Stoetzer LF (2020) How parties react to voter transitions. *Am Polit Sci Rev* 114(3):940–945
- Andreadis I, Chadjipadelis T (2009) A method for the estimation of voter transition rates. *J Elect Public Opin Parties* 19(2):203–218
- Antweiler W (2007) Estimating voter migration in Canada using generalized maximum entropy. *Elect Stud* 26(4):756–771
- Brown PJ, Payne CD (1986) Aggregate data, ecological regression and voting transitions. *J Am Stat Assoc* 81:453–460
- Cavari A, Freedman G (2022) Survey nonresponse and mass polarization: the consequences of declining contact and cooperation rates. *Am Polit Sci Rev*
- Forcina A, Gnaldi M, Bracaleone B (2012) A revised Brown and Payne model of voting behaviour applied to the 2009 elections in Italy. *Stat Methods Appl* 21:109–119
- Füle E (1994) Estimating voter transitions by ecological regression. *Elect Stud* 13:313–330
- Greiner DJ (2007) Ecological inference in voting rights act disputes: where are we now, and where do we want to be? *Jurimetrics* 47:115–167
- Greiner DJ, Quinn KM (2010) Exit polling and racial bloc voting: combining individual level and RxC ecological data. *Annals Appl Stat* 4:1774–1796
- Groves RM, Dillman DA, Eltinge JL, Little RJA (2002) *Survey nonresponse*. Wiley, New York
- Hawkes AG (1969) An approach to the analysis of electoral swing. *J R Stat Soc Ser A* 132:68–79
- Irwin GA, Meeter DA (1969) Building voter transition models from aggregate data. *Midwest J Polit Sci* 13(4):545–566
- Johnston RJ, Pattie CJ (1991) Evaluating the use of entropy-maximising procedures in the study of voting patterns: Sampling and measurement error in the flow-of-the-vote matrix and the robustness of estimates. *Environ Plan A* 23(3):411–420
- Johnston RJ, Pattie CJ (1993) Entropy-maximizing and the iterative proportional fitting procedure. *Prof Geogr* 45(3):317–322
- Klein JM (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 PW, Molnar C, Schlesinger T, Küchenhoff H (2016) Estimation of voter transitions based on ecological inference: an empirical assessment of different approaches. *AStA* 100:133–159
- Klima A, Schlesinger T, Thurner PW, Küchenhoff H (2019) Combining aggregate data and exit polls for the estimation of voter transitions. *Sociol Methods Res* 48:296–325
- McCarthy C, Terence MR (1977) Estimates of voter transition probabilities from the British General Elections of 1974. *J R Stat Soc Ser A* 140:78–85
- Park W-H (2008) Ecological inference and aggregate analysis of elections. PhD Dissertation. The University of Michigan.
- Park W-H, Hanmer MJ, Biggers DR (2014) Ecological inference under unfavorable conditions: straight and split-ticket voting in diverse settings and small samples. *Elect Stud* 36:192–203
- Pavía JM (2022) ei.Datasets: real datasets for assessing ecological inference algorithms. *Soc Sci Comput Rev* 40:247–260
- Pavía JM, Aybar C (2020) Electoral mobility in the 2019 elections in the Valencian region. *Debats. J Cult Power Soc* 134(1):27–51.
- Pavía JM, Romero R (2022a) Improving estimates accuracy of voter transitions. Two new algorithms for ecological inference based on linear programming. *Sociol Methods Res*. Online available.

- Pavía JM, Romero R (2022b) Iphom: ecological Inference by Linear Programming under Homogeneity. R package version 0.3.1–1. <https://CRAN.R-project.org/package=iphom>
- Pavía JM, Romero R (2022c) Symmetry estimating R×C ecological tables vote transfer matrices from aggregate data. Under review.
- Pavía JM, Veres-Ferrer EJ (2016a) Un nuevo estimador para disgregar totales poblacionales. El caso de los nuevos electores. *Anales de Economía Aplicada* 817–826.
- Pavía JM, Veres-Ferrer EJ (2016b) Desagregando Estadísticas de Población. In: Herrerías JM, Callejón J (eds.) *Investigaciones en Métodos Cuantitativos para la Economía y la Empresa*, Editorial Universidad de Granada, pp 543–555.
- Pavía JM, Badal E, García-Cárceles B (2016) Spanish exit polls: sampling error or nonresponse bias? *Rev Int Sociol* 74(3):e043
- Plescia C, De Sio L (2018) An evaluation of the performance and suitability of RxC methods for ecological inference with known true values. *Qual Quant* 52:669–683
- Puig X, Ginebra J (2014) A Bayesian cluster analysis of election results. *J Appl Stat* 41:73–94
- Puig X, Ginebra J (2015) Ecological inference and spatial variation of individual behavior: national divide and elections in Catalonia. *Geogr Anal* 47(3):262–283
- Research Association New Zealand (2020) New Zealand Political Polling Code. www.researchassociation.org.nz, June 2020.
- Romero R, Pavía JM, Martín J, Romero G (2020) Assessing uncertainty of voter transitions estimated from aggregated data. Application to the 2017 French presidential election. *J Appl Stat* 47(13–15):2711–2736
- Russo L (2014) Estimating floating voters: a comparison between the ecological inference and the survey methods. *Qual Quant* 48:1667–1683
- Sandoval P, Ojeda S (2022) Estimation of electoral volatility parameters employing ecological inference methods. *Qual Quant*
- Thomsen SR (1987) Danish elections, 1920–79: a logit approach to ecological analysis and inference. Aarhus: Politica.
- Turner PW, Mauerer I, Bort M, Klima A, Küchenhoff H (2020) Integrating large-scale online surveys and aggregate data at the constituency level: the estimation of voter transitions in the 2015 British General Elections. *Survey Res Methods* 14(5):461–476
- Turner PW, Klima A, Küchenhoff H, Mauerer I, Mang S, Walter-Rogg M, Heinrich T, Knieper T, Schnurbus J (2022) Micromotives of vote switchers and macrotransitions: the case of the immigration issue in a regional earthquake election in Germany 2018. *Politische Vierteljahresschrift* 63:663–684
- Tziafetas G (1986) Estimation of the voter transition matrix. *Optimization* 17:275–279
- Upton CJG (1978) A note on the estimation of voter transition probabilities. *J R Stat Soc Ser A* 141:507–512
- Vangrevelinghe G (1961) Étude statistique comparée des résultats des référendums de 1958 et 1961. *Revue De Statistique Applique* 9:83–100
- Vizcaino A, Pavía JM (2022) New parties and matryoshka scissions in Spain. The case of Podemos and Más Madrid. *Social Sciences & Humanities Open* 6:100307
- Wellhofer ES (2001) Party realignment and voter transitions in Italy, 1987–1996. *Comp Pol Stud* 34(2):156–186