



# **“Ecological Inference: A Minimum Distance to First Moment Constraints Approach”**

**TESIS PARA OPTAR AL GRADO DE  
MAGÍSTER EN ECONOMÍA**

**Alumno: Benjamín Peña Sotomayor  
Profesores Guía: Eduardo Engel y Juan Díaz**

**Santiago, agosto de 2023**

# Ecological Inference: A Minimum Distance to First Moment Constraints Approach\*

Benjamín Peña<sup>†</sup>

Universidad de Chile

August 17, 2023

## Abstract

This paper presents a new approach to the ecological inference problem, particularly to estimating voter transition matrices between two elections. Our estimators choose the points that most closely conform to constraints derived from several first moment conditions arising from mild and natural assumptions. We show that under these assumptions our estimators are consistent, that our estimation procedure has properties that simplify their computation significantly, and derive estimators for the standard deviations of the true voter transitions. We also show that our estimators perform well in small samples through a simulation study and illustrate our approach with three applications. The first application is the well-known problem in the ecological inference literature of estimating the fraction of voter registrations among different demographic groups in the US. The second application uses data on the 2013 Chilean presidential election to analyze voter turnout between the first round and the runoff. We use this application to show that, unlike more computationally intensive approaches, our model can use large datasets without issue. We also show that our approach may be extended in a straightforward manner to the case of multiple clusters in the true values for the voter transitions across units. Both of the previous applications have known true voter transition matrices, and we find that our model performs similarly to more established approaches and provides superior estimates in some circumstances. Our final application estimates voter transitions between the 2021 Chilean presidential election runoff and the 2022 Constitutional Plebiscite, where voters decided whether to approve or reject the proposed draft for a new constitution. Our results for this application suggest that the compulsory voting policy put in place for the 2022 plebiscite significantly impacted its outcome. We find that close to 90% of voters who did not vote in the 2021 runoff but did vote in the 2022 plebiscite voted to reject the constitutional draft. Under the assumption that the bulk of these voters would not have voted had there not been a compulsory voting policy in place, in its absence the result would likely have been the approval of the draft rather than its rejection. Overall, our approach is a viable alternative to other ecological inference methods. It has good theoretical properties and a good performance in small samples, and it is simpler and less computationally intensive than prevailing simulation-based strategies.

---

\*Thesis: M.A. in Economics at Universidad de Chile. I thank Miguel Jorquera for providing the data for one of the applications and Eduardo Engel and Juan Díaz for their guidance in this research project.

<sup>†</sup>bpenas@fen.uchile.cl

# 1. Introduction

Ecological inference is the process of conducting inference at the individual level from aggregate data. The fundamental challenge in doing so is that many different individual behaviors can produce the same observation at the aggregate level. To capture the previous idea, the seminal work of [Robinson \(1950\)](#) first formalized the concept of the ecological fallacy, i.e., the wrongful assumption that aggregate-level correlations are a good substitute for individual-level relations. This resulted in a vast literature, first in sociology and then statistics and political science, that has attempted to find improved methodological strategies for the ecological inference problem.

An important application of ecological inference is the estimation of voter transition matrices, which give us information on how groups of voters move between different candidates or competing options across elections. As an example, consider two consecutive elections where, in each, we group voters by their candidate of choice. A voter transition matrix between both elections would then contain the fraction of voters for each candidate in the first election that votes for each candidate in the second election<sup>1</sup>. An ecological inference problem arises in this context since elections usually yield data aggregated at the level of some unit, such as ballot boxes, rather than data for each voter. Therefore, in most applications, voter transition matrices must be estimated using only aggregate data.

Voter transition matrices are relevant both for the analysis of elections and for the study of voter behavior, and their estimation has a long history in political science and statistics (see, for example, [McCarthy and Ryan 1977](#); [Upton 1978](#); [Brown and Payne 1986](#); [Füle 1994](#); [King et al. 2008](#)). They also have applications in economics<sup>2</sup>. For example, voter transition matrices and ecological inference tools have been employed in political science to study expressive and strategic voting (see [Núñez 2016](#)), an essential subject within political economy<sup>3</sup>.

This paper presents a novel approach to ecological inference, contributing to the broad methodological literature that studies this problem in the social sciences. We focus on the problem of ecological inference applied to estimating voter transition matrices, thereby

---

<sup>1</sup>Voter transition matrices may also be applied in other ways. For example, consider grouping voters in an election into different demographic groups, such as men and women or white and colored voters. A voter transition matrix for this election would give us the fraction of voters from each demographic that votes for each available candidate or option.

<sup>2</sup>For an example of the use of ecological inference in the economics literature outside of political economy, see [Manski \(2018\)](#), who analyzes patient care in an ecological inference framework.

<sup>3</sup>Several approaches other than ecological inference have been taken to study the importance of strategic voting in elections in the economics literature, such as regression discontinuity designs ([Pons and Tricaud, 2018](#)) and structural approaches ([Kawai and Watanabe, 2013](#)).

contributing to the literature on this issue as well. Our approach is simple and has a direct interpretation. Its estimates of the main values of interest, the values composing the voter transition matrices, consist of the points in the unit square that minimize the squared Euclidean distance to several constraints derived from data-imposed first moment conditions. These first moment conditions, on the other hand, arise from mild and natural assumptions. We do not make strong assumptions about the distributions of the parameters in our model, meaning our method may also be described as non-parametric.

Our approach has several properties that make it appealing. Under our base assumptions, our estimators are consistent. Furthermore, they are computed by solving a convex optimization problem with a strictly convex and quadratic objective function, simplifying estimation by eliminating the concern of local optima and reducing its computational intensity compared to simulation-based ecological inference approaches. This makes our method particularly attractive for problems where we are interested in voter transition matrices with many groups of voters in each election and when we have data for a large number of units. Moreover, it provides estimates for the unit-level voter transitions and for the variances of these voter transitions, and it can be extended in a simple way to deal with the presence of multiple clusters in the true unit-level voter transitions, which is something that may arise in certain applications. Finally, it performs well in small samples, which we show through a simulation study.

We illustrate our approach with three applications. In two of these applications, besides the aggregate data, we have data on the true values for the voter transition matrices, which allows us to directly assess the performance of our model and compare it to select alternative approaches.

In our first application, we estimate the fraction of the white and colored voting-age population who register to vote in four states in the US using a classic dataset in the ecological inference literature (see [King 1997](#); [King, Rosen, and Tanner 1999](#); [King, Tanner, and Rosen 2004](#)). We find that our model performs similarly to more standard approaches in estimating the true voter transition matrix.

Our second application uses data on voter turnout during the first round and runoff of the 2013 Chilean presidential election. We group the voting-age population into voters and non-voters in each election and estimate the fraction of first-round voters and non-voters that vote in the runoff. In this case, our model outperforms other approaches. This application also shows that our approach can be used to conduct inference with large datasets where more computationally intensive methods become unviable and that a simple, direct extension of our base estimation strategy may be employed to take advantage

of the presence of multiple clusters to improve its efficiency.

Our third and final application uses our model to estimate the voter transition matrix between the 2021 Chilean presidential election runoff and the 2022 Chilean Constitutional Plebiscite, where voters decided whether to approve or reject the proposed draft for a new constitution. Despite not having data on the true voter transition matrix, this application is interesting due to the social and political relevance of both elections. We arrive at several notable results. In particular, we find that the compulsory voting policy put in place for the 2022 plebiscite likely had a significant impact on its outcome. Voters who had not voted in the presidential election runoff but did vote in the 2022 plebiscite overwhelmingly favored the option to Reject. Concretely, 88% of these voters opted to reject the draft. Under the assumption that most of these voters would have chosen not to vote in the 2022 plebiscite under a voluntary voting policy, our results suggest that the compulsory voting policy put in place for this vote effectively flipped the result from a likely approval of the draft to its resounding rejection.

Overall, our approach is viable as an alternative and a complement to other ecological inference approaches. It has good theoretical properties and a good performance in small samples, and it is simpler and less computationally intensive than prevailing simulation-based strategies.

The rest of the paper is organized as follows. Section 2 describes in detail the ecological inference problem in the context of estimating voter transition matrices. Section 3 briefly reviews some of the most relevant literature on ecological inference, particularly the literature on ecological inference in the social sciences and statistics. Section 4 presents our model and proposed estimators for all parameters, proving their consistency and several other properties under mild assumptions. Section 5 presents a simulation study where we analyze the performance of our estimators in small samples under controlled conditions. Section 6 illustrates our estimation strategy using two ecological inference problems with known true voter transition matrices. The first is the estimation of voter registration rates for white and colored voters in the US, while our second application analyzes voter turnout in both rounds of the 2013 Chilean presidential elections. Section 7 presents the application to the 2021 Chilean presidential election runoff and the 2022 Constitutional Plebiscite. Section 8 concludes.

## 2. The Ecological Inference Problem

We take the estimation of voter transition matrices between two elections as a running example. The simplest case of ecological inference is known as the 2x2 case, in which we divide the voting-age population (in what follows, voters) into two groups in each election. The objective is then to learn about the voter transitions across both elections for the two groups of voters in the first election. As was previously mentioned, what makes this an ecological inference problem is that we only observe the total number of votes for each group in each election aggregated at the level of some unit.

As a concrete example of a 2x2 ecological inference problem, consider an application to voter turnout where we divide voters in each election into voters and non-voters. In this case, we are interested in the fraction of first-election voters and non-voters that vote in the second election. These two fractions describe the voter transition matrix between these elections for our defined groups of voters.

To formalize the previous example, assume that the available data is aggregated at the level of some unit (e.g., ballot boxes) that we index by  $i$ , where  $i = 1, 2, \dots, N$ . For each unit, we have data on  $n_i$ , the number of votes in the first election,  $m_i$ , the number of non-voters in the first election (so that  $n_i + m_i$  is the total number of voters in unit  $i$ ) and  $v_i$ , the number of votes in the second election. We assume that the total number of voters does not change between elections. Let  $p_i$  and  $q_i$  be the fraction of first-election voters and non-voters that vote in the second election. These values compose the unit-level voter transition matrix for unit  $i$ . Since only aggregate data is available, both are unknown. However, we can identify some relations between these unknown values and our available data. Table 1 illustrates these relations for each unit.

Table 1: Ecological Inference - 2x2 Case

|                |           | Second Election     |  |       |
|----------------|-----------|---------------------|--|-------|
| First Election | Voters    | Non-Voters          |  |       |
| Voters         | $n_i p_i$ | $n_i(1 - p_i)$      |  | $n_i$ |
| Non-Voters     | $m_i q_i$ | $m_i(1 - q_i)$      |  | $m_i$ |
|                | $v_i$     | $(n_i + m_i) - v_i$ |  |       |

*Notes:* This table illustrates the ecological inference problem in the simplest possible scenario. The available data consists of  $n_i$ ,  $m_i$ , and  $v_i$ , which in this example are the total number of voters in the first election, the total number of non-voters in the first election, and the total number of voters in the second election, respectively. The unknown variables are  $p_i$  and  $q_i$ , the fraction of first-election voters and non-voters who vote in the second election in unit  $i$ . Each entry at the center of the table in row  $r$  and column  $c$  indicates the number of voters from the group in row  $r$  that also belong to the group in column  $c$ .

Each entry at the center of the table in row  $r$  and column  $c$  indicates the number of voters from the group in row  $r$  that belong to the group in column  $c$ . For example, the value at the second row and first column,  $m_i q_i$ , corresponds to the number of voters that did not vote in the first election but did vote in the second. Each unit has a table such as this, where the values at the center are unknown and the only observable values are the aggregate quantities at the margins of the table.

In a typical ecological inference application, we are particularly interested in the aggregate (across all units) fraction of first-election voters and non-voters that vote in the second election, which we denote by  $\mu_p$  and  $\mu_q$ , respectively. These values compose the *aggregate* voter transition matrix between the two elections. Since these values depend on each  $p_i$  and  $q_i$ , they are also unknown. Ecological inference techniques aim to use the information in tables such as the above to learn about  $p_i$ ,  $q_i$ ,  $\mu_p$ , and  $\mu_q$ .

The main challenge in an ecological inference problem such as this is that multiple combinations of  $p_i$  and  $q_i$  are compatible with the data at the margins of Table 1. In particular, every pair  $p_i, q_i$  for which both values lie between 0 and 1 and such that the condition

$$n_i p_i + m_i q_i = v_i \tag{1}$$

holds is compatible with our available information. Equations such as (1) are known as tomography lines in the ecological inference literature. The fact that infinite values of  $p_i$  and  $q_i$  are compatible with Table 1 means that we can only obtain limited insights on the unit-level fractions from individual tables such as the above. The main idea behind most ecological inference methods, therefore, is to combine the information of multiple tables with assumptions about the distribution of  $p_i$  and  $q_i$  across units to conduct inference on the unknown fractions.

The representation of the problem in Table 1 lends itself to a direct extension to the more general case, known as the RxC case. In this case, we divide voters in the first election into  $R$  groups and voters in the second election into  $C$  groups by their candidates of choice in each. We are interested in the fraction of voters for each candidate in the first election that votes for each candidate in the second. As before, these fractions compose the voter transition matrix between these elections, and we may be particularly interested in unit-level fractions or aggregate fractions across all units depending on the application. As in the 2x2 case, only the total number of votes for each group in each election are available for each unit  $i$ , so all previous values are unknown. Table 2 illustrates this in an analogous manner to Table 1, where  $n_{r_i}$  is the number of votes for option  $r = 1, \dots, R$  in unit  $i$  in the first election,  $v_{c_i}$  is the number of votes for option  $c = 1, \dots, C$  in unit  $i$  in the

second election, and  $p_{rci}$  is the fraction of voters for option  $r$  in the first election who vote for  $c$  in the second election in unit  $i$ .

Table 2: Ecological Inference - RxC Case

| First Election      | Second Election     |                     |     |                     |          |
|---------------------|---------------------|---------------------|-----|---------------------|----------|
|                     | Voters for option 1 | Voters for option 2 | ... | Voters for option C |          |
| Voters for option 1 | $n_{1i}p_{11i}$     | $n_{1i}p_{12i}$     | ... | $n_{1i}p_{1Ci}$     | $n_{1i}$ |
| Voters for option 2 | $n_{2i}p_{21i}$     | $n_{2i}p_{22i}$     | ... | $n_{2i}p_{2Ci}$     | $n_{2i}$ |
| ⋮                   | ⋮                   | ⋮                   | ⋮   | ⋮                   | ⋮        |
| Voters for option R | $n_{Ri}p_{R1i}$     | $n_{Ri}p_{R2i}$     | ... | $n_{Ri}p_{RCi}$     | $n_{Ri}$ |
|                     | $v_{1i}$            | $v_{2i}$            | ... | $v_{Ci}$            |          |

*Notes:* This table illustrates the ecological inference problem in the general scenario with  $R$  groups in the first election and  $C$  groups in the second election. The observed data consists of  $n_{ri}$ , the number of voters in group  $r = 1, 2, \dots, R$  in unit  $i$  in the first election, and  $v_{ci}$ , the number of voters in group  $c = 1, \dots, C$  in unit  $i$  in the second election. The unknown variables are  $p_{rci}$  for each  $r, c$ , and  $i$ , which indicates the fraction of people that vote for option  $r$  who also vote for option  $c$  in unit  $i$ . Each entry at the center of the table in row  $r$  and column  $c$  indicates the number of voters from the group in row  $r$  that belong to the group in column  $c$ .

Similarly to Table 1, the information that we may gather from Table 2 is that  $p_{rci} \in [0, 1]$  for all  $r, c$ , and  $i$ , that  $\sum_{c=1}^C p_{rci} = 1$  for each  $r$  and for each unit  $i$ , since each voter votes for one and only one candidate in the second election, and that the following relations hold in each unit:

$$\begin{aligned} \sum_{r=1}^R n_{ri}p_{r1i} &= v_{1i} \\ \sum_{r=1}^R n_{ri}p_{r2i} &= v_{2i} \\ &\vdots \\ \sum_{r=1}^R n_{ri}p_{rCi} &= v_{Ci}, \end{aligned}$$

Infinite values for the fractions  $p_{rci}$  are compatible with the previous conditions. Therefore, as was the case for 2x2 ecological inference, we have very limited information about these fractions from one unit alone.

Having described the problem in sufficient detail, we now review the most relevant literature on the subject.

### 3. Literature Review

The seminal paper on ecological inference is [Robinson \(1950\)](#), which shows that ecological correlation, or correlation using aggregate quantities, is not always a good substitute for correlation at the individual level. Following Robinson’s findings, [Duncan and Davis \(1953\)](#) and [Goodman \(1953, 1959\)](#) proposed alternatives to ecological correlation for the 2x2 case. Respectively, they proposed the deterministic method of bounds, which involves using the data of multiple tables such as [Table 1](#) and [Table 2](#) to infer lower and upper bounds on the quantities of interest, and ecological regressions, which are simply regressions involving the aggregate quantities across units. Due to this contribution, regressions involving aggregate quantities have come to be known as Goodman regressions in the literature.

Both of these approaches had several issues, however. The former is limited in its ability to provide insight into the values we are interested in, given that, in general, the bounds determined by each table are not narrow enough to provide valuable information. On the other hand, Goodman regressions will only work correctly under strong assumptions, such as the assumption of homogenous voter transitions across all units. Despite their drawbacks, most of the literature following these papers focused on expanding upon these approaches or on using them in different applications.

The next major contribution to the literature was [King \(1997\)](#), which presents a model combining insights from both previous approaches into a simulation-based approach with a better performance than either<sup>4</sup>. Building on the insight of [King \(1997\)](#), [King, Rosen, and Tanner \(1999\)](#) then developed a Bayesian hierarchical model using a Binomial-Beta conjugate in the first two levels and an exponential hyperprior for the parameters of the Beta distributions. Some contributions to the literature on Bayesian approaches to 2x2 ecological inference since then include [Wakefield \(2004\)](#) and [Imai, Lu, and Strauss \(2008\)](#).

Despite the significant contribution of the [King, Rosen, and Tanner \(1999\)](#) model to the ecological inference literature, among which is the fact that it is easily extensible to the RxC case, it also has several drawbacks. Since Bayesian methods result in distributions for the quantities of interest rather than point estimates, estimation is carried out by sampling from the posterior distribution of the model using Markov-Chain Monte Carlo (MCMC) methods. When coupled with their complex nature, this makes hierarchical models have a high computational cost. Hierarchical models also have some other issues, such as being somewhat of a “black box” and being sensible to the choice of initial parameter values. Another challenge with this approach, which stems from the simulation of unknown

---

<sup>4</sup>For a critical view of [King \(1997\)](#), see [Freedman et al. \(1998\)](#), [King \(1999\)](#), and [Freedman et al. \(1999\)](#).

distributions via MCMC, is that it is difficult to know when the chain has converged. This is especially problematic given that these models lose their value unless we draw samples from the correct posterior distribution.

Following [King, Rosen, and Tanner \(1999\)](#), [Rosen et al. \(2001\)](#) formalized an RxC Bayesian hierarchical model with a Dirichlet-Multinomial conjugate in its first two levels based directly on the 2x2 model. Like the [King, Rosen, and Tanner \(1999\)](#) model, inference is carried out via MCMC sampling. As in the 2x2 case, the need to draw from a posterior distribution makes this model quite computationally intensive. It also shares the issues with hierarchical models that were outlined above. Among recent methodological contributions in this line of research, [Greiner and Quinn \(2009\)](#) provides an alternative hierarchical model using different distributions, while [Klima et al. \(2019\)](#) expands the [Rosen et al. \(2001\)](#) model to supplement it with survey data. Alternative approaches include [Andreadis and Chadjipadelis \(2009\)](#), which presents an alternative model approaching the RxC case by splitting it into multiple, sequential 2x2 problems, and [Pavía and Romero \(2022\)](#), which views the ecological inference problem from a linear programming perspective rather than as a statistical inference problem and presents an algorithm for estimating transition matrices. Importantly, in independent evaluations of different ecological inference models [Klima et al. \(2016\)](#) and [Plescia and De Sio \(2018\)](#) both find that a particular implementation of the Bayesian hierarchical model in [Rosen et al. \(2001\)](#) through the eiPack library in R<sup>5</sup> ([Lau, Moore, and Kellermann, 2007](#)) is generally superior to the alternatives they consider.

Of the work that is not directly a part of the ecological inference literature, this paper is also related to [Zubizarreta \(2015\)](#) and [Chattopadhyay and Zubizarreta \(2022\)](#), which propose and analyze the properties of estimators that are related to our approach but are applied to the issue of causal inference with non-experimental data.

## 4. A Minimum Distance to First Moment Constraints Approach

We present our model for the 2x2 case and briefly state the extension to the RxC case at the end of this section. Going back to the example of two consecutive elections and voter turnout in [Section 2](#), we now define  $p_i$  as the *probability* that first-election voters vote again in the second election and  $q_i$  as the probability that first-election non-voters vote in the second election. We assume that these probabilities across units are realizations of independent and identically distributed (i.i.d.) random variables with means  $\mu_p$  and  $\mu_q$

---

<sup>5</sup>This is important given that eiPack implements the [Rosen et al. \(2001\)](#) model using a gamma distribution for the hyperprior rather than an exponential distribution, which is the originally proposed distribution for the hyperprior.

and variances  $\sigma_p^2$  and  $\sigma_q^2$ , respectively, but do not specify their distribution. As in Section 2, for each unit  $i$ , let  $n_i$  be the number of votes in the first election and  $m_i$  the number of non-voters in the first election, and assume that the total voting age population (in what follows, voters) does not change between elections. Finally, assume that voters vote independently from each other. This implies that the number of votes in the second round in each unit,  $v_i$ , is the sum of two independent Binomial random variables, one with  $n_i$  trials and probability  $p_i$ , corresponding to the number of votes from first-round voters, and the other with  $m_i$  trials and probability  $q_i$ , corresponding to the number of votes from first-round non-voters. We summarize all the previous assumptions in Assumption 1 for future reference.

**Assumption 1.** *The distribution of the unit-level parameters and the number of votes in each unit  $i$  satisfies the following:*

$$\begin{aligned} p_i &\sim (\mu_p, \sigma_p^2) \\ q_i &\sim (\mu_q, \sigma_q^2) \\ v_i &= \text{Bin}(n_i, p_i) \oplus \text{Bin}(m_i, q_i) \end{aligned}$$

where  $\oplus$  denotes the sum of independent random variables. Furthermore,  $p_i$  and  $q_i$  are *i.i.d.* random variables across units, and  $p_i$  is independent of  $q_i$  within units.

Note that at no point in Assumption 1 do we make strong distributional assumptions, since  $v_i$  being the sum of two independent Binomial random variables is a consequence of assuming that voters vote independently. This implies that our approach may also be described as non-parametric, although we do not emphasize this view in this paper.

Assumption 1 implies that if we plot the realized parameters  $(p_i, q_i)$  in the Cartesian plane with each  $p_i$  in the x-axis and each  $q_i$  in the y-axis, we will (in the general case) see a single cluster of points<sup>6</sup>. This is similar to other common assumptions in the literature on 2x2 ecological inference (King, 1997; King, Rosen, and Tanner, 1999; Wakefield, 2004). For instance, the King, Rosen, and Tanner (1999) model assumes that the prior distributions of the unit-level fractions (which are analogous to our probabilities  $p_i$  and  $q_i$ ) are independent Beta distributions. Note, however, that we do not make the assumption of a single cluster in the voter probabilities.

---

<sup>6</sup>There may be two clusters in the case of a bimodal distribution. We consider this case in one of the applications in Section 4.

It also follows from Assumption 1 that

$$\mathbb{E}(v_i|p_i, q_i) = n_i p_i + m_i q_i, \quad (2)$$

which is analogous to (1), the condition for the unit-level fractions to be consistent with the data discussed in Section 2. However, note that (2) is a statement on a conditional expectation involving *unknown parameters*  $p_i$  and  $q_i$ . In contrast, (1) involves the *realized fractions* of voters and non-voters in the first election that vote in the second election.

Our main goal is to estimate  $\mu_p$  and  $\mu_q$ . We begin by noting that if the unit-level parameters  $p_i$  and  $q_i$  have a common mean, we can attempt to estimate  $\mu_p$  and  $\mu_q$  by choosing the point  $(\hat{p}, \hat{q})$  in the unit square that is closest to the cluster (or clusters) of points  $p_i$  and  $q_i$ . Since  $p_i$  and  $q_i$  are unobservable, and in the absence of voter-level data, one way to measure the position of the unit-level parameters is to use data-imposed restrictions on the values of  $p_i$  and  $q_i$ . In particular, and in the spirit of the method of moments, we can use the first moment condition (2) to arrive at a constraint involving  $p_i$ ,  $q_i$ , and observable quantities by replacing  $\mathbb{E}(v_i|p_i, q_i)$  with its sample mean. Since we only have one observation for each unit, this sample mean is  $v_i$ . This leads to the following first moment constraint on the values of  $p_i$  and  $q_i$  for each unit:

$$v_i = n_i p_i + m_i q_i. \quad (3)$$

With (3), each unit defines a line in the Cartesian plane with  $p_i$  on the x-axis and  $q_i$  on the y-axis on which we are restricting our estimates of  $p_i$  and  $q_i$  to lie. Note, however, that the actual unit-level parameters do not have to lie in (3) to be consistent with the values of  $v_i$ ,  $n_i$ , and  $m_i$  in each unit. The parameters  $p_i$  and  $q_i$  determine the realized fraction of voters from each group in the first election (voters and non-voters) that vote in the second election. As discussed in Section 2, it is these *fractions* that have to be consistent with the data in the manner of (1) and (3). The main point is that, given the available data, deriving constraint (3) from condition (2) provides us with the best available measure of where the pair  $(p_i, q_i)$  has to lie in the unit square for each unit.

Having a way to locate the unit-level parameters, we can estimate  $\mu_p$  and  $\mu_q$  by choosing the point in the unit square that minimizes the sum of the squared distances to each of these lines. We can then obtain estimates for the unit-level parameters by choosing the points in the unit square on each line as defined in (3) closest to  $(\hat{p}, \hat{q})$ . Due to this intuition for our model, we refer to it as the “Minimum Distance” model in what follows. Figure 1 shows how our estimator works in a simple scenario with six units, where the red

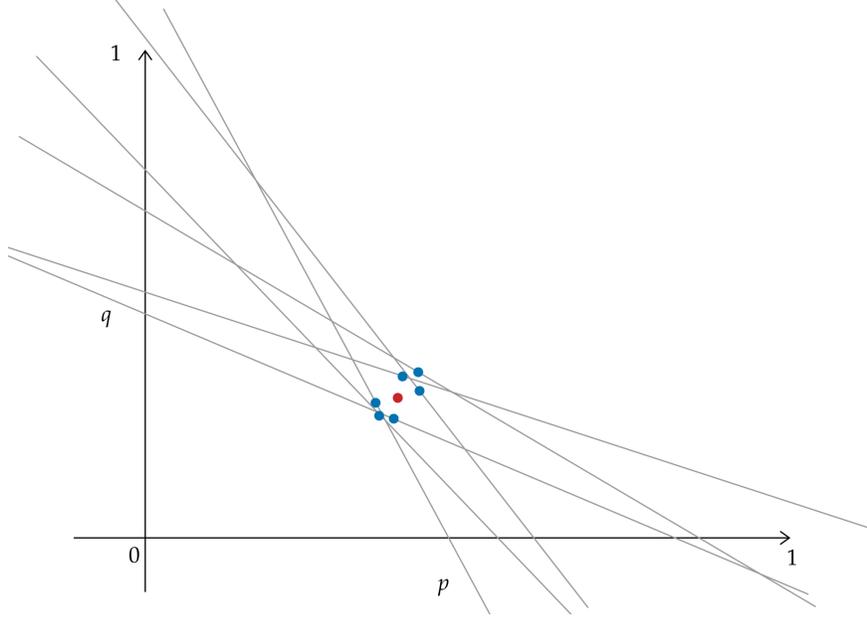


Figure 1: Estimation - Simple Example

*Notes:* This figure presents a simple example of our estimation procedure. Each line corresponds to a first moment condition defined by one unit. We use these to impose constraints on the values of  $p_i$  and  $q_i$ . The red dot corresponds to the pair  $(\hat{p}, \hat{q})$ , our estimators for  $(\mu_p, \mu_q)$ . This is the point that minimizes the sum of the distances to each of the lines in the unit square. The blue dots correspond to the pairs  $(\hat{p}_i, \hat{q}_i)$ , which are the closest points on each line to  $(\hat{p}, \hat{q})$  in the unit square.

dot corresponds to the  $(\hat{p}, \hat{q})$  pair and the blue dots to each  $(\hat{p}_i, \hat{q}_i)$  pair.

Each line corresponds to the line defined by (3) for each unit. Our estimates for  $\mu_p$  and  $\mu_q$ ,  $\hat{p}$  and  $\hat{q}$ , minimize the sum of the squared distances between this pair of points and each line in the unit square, and our estimates for  $p_i$  and  $q_i$  for each unit are the closest points in each line to the  $(\hat{p}, \hat{q})$  pair. Formally, this problem may be stated as a bi-level optimization problem:

$$\begin{aligned}
 \min_{\{\hat{p}, \hat{q}\} \in [0,1]^2} \quad & \sum_{i=1}^N [(\hat{p}_i(\hat{p}, \hat{q}) - \hat{p})^2 + (\hat{q}_i(\hat{p}, \hat{q}) - \hat{q})^2] \\
 \text{s.t.} \quad & (\hat{p}_i(\hat{p}, \hat{q}), \hat{q}_i(\hat{p}, \hat{q})) \in \underset{\hat{p}_i, \hat{q}_i \in [0,1]}{\operatorname{argmin}} [(\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2], \quad i = 1, \dots, N \\
 & \text{s.t. } v_i = n_i \hat{p}_i + m_i \hat{q}_i
 \end{aligned} \tag{4}$$

In terms of Figure 1, the inner level of the problem, which is the restriction to the outer level, finds the values for  $\hat{p}_i, \hat{q}_i$  in the unit square in each line closest to a given  $\hat{p}$  and  $\hat{q}$ . The outer level of the problem then chooses the values for  $\hat{p}$  and  $\hat{q}$  in the unit square for which the sum of the squared distances between  $(\hat{p}, \hat{q})$  and each of the previously defined

points is minimized.

If we ignore the bound constraints or assume that the optimal  $\hat{p}$  and  $\hat{q}$  fall between 0 and 1 and that, at these points, all  $(\hat{p}_i(\hat{p}, \hat{q}), \hat{q}_i(\hat{p}, \hat{q}))$  likewise belong to the unit square, we can obtain a system of equations for  $\hat{p}$  and  $\hat{q}$  by solving a series of unconstrained optimization problems. First, we solve for the individual level fractions for given  $\hat{p}$  and  $\hat{q}$  at the inner level in (4) to arrive at the following solution for each  $\hat{p}_i$  and  $\hat{q}_i$  for each unit:

$$\begin{aligned}\hat{p}_i(\hat{p}, \hat{q}) &= \frac{n_i}{n_i^2 + m_i^2} v_i + \frac{m_i^2}{n_i^2 + m_i^2} \hat{p} - \frac{n_i m_i}{n_i^2 + m_i^2} \hat{q} \\ \hat{q}_i(\hat{p}, \hat{q}) &= \frac{m_i}{n_i^2 + m_i^2} v_i + \frac{n_i^2}{n_i^2 + m_i^2} \hat{q} - \frac{n_i m_i}{n_i^2 + m_i^2} \hat{p}\end{aligned}\tag{5}$$

We then substitute these expressions back into the outer level of (4) and solve the minimization problem to arrive at the following system of equations:

$$\begin{bmatrix} \hat{p} \\ \hat{q} \end{bmatrix} = N \begin{bmatrix} \sum_{i=1}^N n_i^2 w_i & \sum_{i=1}^N n_i m_i w_i \\ \sum_{i=1}^N n_i m_i w_i & \sum_{i=1}^N m_i^2 w_i \end{bmatrix}^{-1} \frac{1}{N} \begin{bmatrix} \sum_{i=1}^N n_i w_i v_i \\ \sum_{i=1}^N m_i w_i v_i \end{bmatrix},\tag{6}$$

where  $w_i = (n_i^2 + m_i^2)^{-1}$ . These expressions will be useful below. We can also obtain closed-form expressions for the pair  $(\hat{p}, \hat{q})$  as well as the unit-level estimates  $(\hat{p}_i, \hat{q}_i)$  by solving (6) and then plugging these solutions into (5), but we gain nothing by doing so here. They are shown in Appendix C.

We now prove several properties of our estimators. First, note that  $\hat{p}$  and  $\hat{q}$  equal the average of the unit level estimates  $\hat{p}_i$  and  $\hat{q}_i$ , respectively. We can see this by taking the average across units of both equations in (5), setting  $\hat{p} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i(\hat{p}, \hat{q})$  and  $\hat{q} = \frac{1}{N} \sum_{i=1}^N \hat{q}_i(\hat{p}, \hat{q})$ , letting  $w_i = (n_i^2 + m_i^2)^{-1}$ , and then solving the resulting system of two equations for  $\hat{p}$  and  $\hat{q}$  to arrive at (6). We can use this in (4) to get a standard (rather than a bi-level) optimization problem that characterizes our estimators. Proposition 1 formalizes this result.

**Proposition 1.** *The optimization problem in (4) is equivalent to the following problem:*

$$\begin{aligned}\min_{\{\hat{p}_i, \hat{q}_i\}_{i=1, \dots, N}} & \sum_{i=1}^N \left[ (\hat{p}_i - \frac{1}{N} \sum_{i=1}^N \hat{p}_i)^2 + (\hat{q}_i - \frac{1}{N} \sum_{i=1}^N \hat{q}_i)^2 \right] \\ \text{s.t.} & \quad n_i \hat{p}_i + m_i \hat{q}_i = v_i, \quad i = 1, \dots, N \\ & \quad \hat{p}_i, \hat{q}_i \in [0, 1], \quad i = 1, \dots, N\end{aligned}\tag{7}$$

*Proof.* See Appendix A □

We can also show that the previous problem has several properties that simplify our estimation procedure. Proposition 2 presents this result.

**Proposition 2.** *The problem presented in (7) is a convex optimization problem. Furthermore, the objective function is strictly convex and quadratic.*

*Proof.* See Appendix B □

This significantly simplifies estimation in two ways. First, since the problem has a unique optimum, local optima are not a concern. Second, we can consider specialized algorithms to speed up estimation.

We now show in the following proposition that the estimators for  $\mu_p$  and  $\mu_q$  in (6) are consistent and asymptotically normal. Note that, in doing this, we are only showing that the estimators without imposing bound constraints have these properties, since (6) was derived under this assumption. Therefore, the consistency and asymptotic normality properties do not necessarily hold for the estimators derived from (4) and (7), since these problems can have active bound constraints. However, doing this is useful in two ways. First, the consistency result allows us to show that the estimators with bound constraints are also consistent. Second, the asymptotic normality result will enable us to arrive at estimators for  $\sigma_p^2$  and  $\sigma_q^2$ . Proposition 3 formalizes the consistency and asymptotic normality properties of the estimators without bound constraints, while the other two results are derived in two corollaries.

**Proposition 3.** *Under Assumption 1 and without imposing bound constraints,  $\hat{p}$  and  $\hat{q}$  as determined by (6) are mean square consistent estimators of  $\mu_p$  and  $\mu_q$ . They are also asymptotically normal with a standard deviation that depends on  $\sigma_p^2$ ,  $\mu_p$ ,  $\sigma_q^2$ , and  $\mu_q$ . In particular, the following holds:*

$$\begin{bmatrix} \hat{p} \\ \hat{q} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mu_p \\ \mu_q \end{bmatrix}$$

$$\sqrt{N} \left( \begin{bmatrix} \hat{p} \\ \hat{q} \end{bmatrix} - \begin{bmatrix} \mu_p \\ \mu_q \end{bmatrix} \right) \xrightarrow{d} \mathcal{N} \left( \mathbf{0}, \text{plim} \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \mathbf{X} \right)^{-1} \text{plim} \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \boldsymbol{\Omega} \mathbf{W} \mathbf{X} \right) \text{plim} \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \mathbf{X} \right)^{-1} \right)$$

where  $\text{plim}$  indicates the limit in probability, and

$$\mathbf{X} = \begin{bmatrix} n_1 & n_2 & \dots & n_N \\ m_1 & m_2 & \dots & m_N \end{bmatrix}'$$

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_N \end{bmatrix}$$

$$\mathbf{\Omega} = \begin{bmatrix} \text{Var}(\varepsilon_1) & 0 & \dots & 0 \\ 0 & \text{Var}(\varepsilon_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Var}(\varepsilon_N) \end{bmatrix}$$

with  $\varepsilon_i = v_i - \mathbb{E}(v_i)$  and  $w_i = [(n_i^2 + m_i^2)]^{-1}$ .

*Proof.* See Appendix C □

As was previously mentioned, Proposition 1 implies that the estimators that result from (4) and (7), where we impose bound constraints on the values of our estimates, are consistent as well. Corollary 1 presents this result.

**Corollary 1.** *Under Assumption 1, the mean-squared error (MSE) of the estimators that result from (4) and (7) is lower or equal to the MSE of the estimators in (6). This implies that the former are consistent estimators of  $\mu_p$  and  $\mu_q$ .*

*Proof.* See Appendix D □

Given that Corollary 1 shows that the constrained estimators are not only consistent but also superior to the unconstrained estimators, we always compute  $\hat{p}$  and  $\hat{q}$  imposing bound constraints. An important point we emphasize here is that the asymptotic normality result of Proposition 3 does not hold when using our estimators with bound constraints. Therefore, we use bootstrap standard errors to obtain their standard deviations when necessary. Furthermore, note that we have only proved the consistency of  $\hat{p}$  and  $\hat{q}$ . This is because although we can obtain estimates for the unit-level parameters, having only one observation for each unit means we cannot say that our estimators  $\hat{p}_i$  and  $\hat{q}_i$  will be

consistent. This is a limitation on all ecological inference models that estimate these or similar unit-level parameters.

We now use the asymptotic normality result of Proposition 3 to derive estimators for  $\sigma_p^2$  and  $\sigma_q^2$ . Corollary 2 presents this result.

**Corollary 2.** *Let  $\hat{\sigma}_p^2$  and  $\hat{\sigma}_q^2$  be defined as follows:*

$$\begin{bmatrix} \hat{\sigma}_p^2 \\ \hat{\sigma}_q^2 \end{bmatrix} = N \begin{bmatrix} \sum n_i^3(n_i - 1)w_i^2 & \sum n_i^2 m_i(m_i - 1)w_i^2 \\ \sum m_i^2 n_i(n_i - 1)w_i^2 & \sum m_i^3(m_i - 1)w_i^2 \end{bmatrix}^{-1} \frac{1}{N} \left\{ \begin{bmatrix} \sum (\hat{p}_i - \hat{p})^2 \\ \sum (\hat{q}_i - \hat{q})^2 \end{bmatrix} - \begin{bmatrix} \sum n_i^3 w_i^2 & \sum n_i^2 m_i w_i^2 \\ \sum n_i m_i^2 w_i^2 & \sum m_i^3 w_i^2 \end{bmatrix} \begin{bmatrix} \hat{p}(1 - \hat{p}) \\ \hat{q}(1 - \hat{q}) \end{bmatrix} \right\}$$

*Under Assumption 1 and if we estimate  $\hat{p}$ ,  $\hat{q}$  and  $\hat{p}_i$  and  $\hat{q}_i$  without bound constraints,  $\hat{\sigma}_p^2$  and  $\hat{\sigma}_q^2$  are consistent estimators for  $\sigma_p^2$  and  $\sigma_q^2$ .*

*Proof.* See Appendix E □

Note that since we rely on the asymptotic normality result of Proposition 3, the estimators in this result are the unconstrained estimators in (5) and (6). Therefore, we estimate  $\sigma_p^2$  and  $\sigma_q^2$  by estimating our model without bound constraints on the value of our estimates and then computing  $\hat{\sigma}_p^2$  and  $\hat{\sigma}_q^2$  in the way described in Corollary 2.

As an alternative, we may also estimate the variance of the unit-level parameters using the sample variance (SV) of our unit-level estimates:

$$\begin{bmatrix} \hat{\sigma}_{p,\text{SV}}^2 \\ \hat{\sigma}_{q,\text{SV}}^2 \end{bmatrix} = \frac{1}{N} \begin{bmatrix} \sum_{i=1}^N (\hat{p}_i - \hat{p})^2 \\ \sum_{i=1}^N (\hat{q}_i - \hat{q})^2 \end{bmatrix}.$$

The rationale behind this is that if the individual estimates and our estimates for  $\mu_p$  and  $\mu_q$  are close to their true values, the sample variances will be close to the variance of the actual parameters across units. Since our estimates for  $p_i$  and  $q_i$  depend on  $v_i$ , however, this is not the case, since the variance of  $v_i$  depends on the underlying variance of  $p_i$  and  $q_i$  across units and the variance of the Binomial random variables. The variance of  $\hat{p}_i$  and  $\hat{q}_i$  across units will be a combination of these two factors. Nevertheless, this is a viable approach if we want an alternative measure of the dispersion of  $p_i$  and  $q_i$ .

So far, we have assumed that all units are equally important, since all of the terms in the objective function in (4) and (7) have the same weight. We now consider the case where we are interested in the total population transition probabilities rather than the expected value of the unit-level parameters. Since different units have different numbers of voters, by the total probability theorem we have that  $\sum_{i=1}^N \psi_i p_i$  and  $\sum_{i=1}^N \psi_i q_i$ , where

$\psi_i = \frac{n_i + m_i}{\sum_{i=1}^N (n_i + m_i)}$ , are the aggregate population transition probabilities across all units, since  $n_i + m_i$  is the total number of voters in each unit in both elections. For  $\hat{p}$  and  $\hat{q}$  to be estimates of these probabilities, we need to weigh units differently in the objective functions of (4) and (7). The following proposition presents this extension and extends the other results in this section to the case with different weights in each unit.

**Proposition 4.** *Let  $\psi_i > 0$ ,  $i = 1, \dots, N$ , with  $\sum_{i=1}^N \psi_i = 1$ . Consider the following generalization of (4):*

$$\begin{aligned} \min_{\{\hat{p}, \hat{q}\}} \quad & \sum_{i=1}^N \psi_i [(\hat{p}_i(\hat{p}, \hat{q}) - \hat{p})^2 + (\hat{q}_i(\hat{p}, \hat{q}) - \hat{q})^2] \\ \text{s.t.} \quad & (\hat{p}_i(\hat{p}, \hat{q}), \hat{q}_i(\hat{p}, \hat{q})) \in \underset{\hat{p}_i, \hat{q}_i \in [0, 1]}{\operatorname{argmin}} [(\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2], \quad i = 1, \dots, N \\ & \text{s.t. } v_i = n_i \hat{p}_i + m_i \hat{q}_i \end{aligned}$$

If we set  $\hat{p} = \sum_{i=1}^N \psi_i \hat{p}_i$  and  $\hat{q} = \sum_{i=1}^N \psi_i \hat{q}_i$ , the previous problem can be rewritten as follows:

$$\begin{aligned} \min_{\{\hat{p}_i, \hat{q}_i\}_{i=1, \dots, N}} \quad & \sum_{i=1}^N \psi_i \left[ (\hat{p}_i - \sum_{i=1}^N \psi_i \hat{p}_i)^2 + (\hat{q}_i - \sum_{i=1}^N \psi_i \hat{q}_i)^2 \right] \\ \text{s.t.} \quad & n_i \hat{p}_i + m_i \hat{q}_i = v_i, \quad i = 1, \dots, N \\ & \hat{p}_i, \hat{q}_i \in [0, 1], \quad i = 1, \dots, N \end{aligned}$$

Ignoring bound constraints or assuming that they are not active, this results in the following consistent estimators for  $\mu_p$  and  $\mu_q$ :

$$\begin{bmatrix} \hat{p} \\ \hat{q} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \psi_i n_i^2 w_i & \sum_{i=1}^N \psi_i n_i m_i w_i \\ \sum_{i=1}^N \psi_i n_i m_i w_i & \sum_{i=1}^N \psi_i m_i^2 w_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N \psi_i n_i w_i v_i \\ \sum_{i=1}^N \psi_i m_i w_i v_i \end{bmatrix}$$

while the expressions for the unit-level estimators  $\hat{p}_i$  and  $\hat{q}_i$  are unchanged. The results of Proposition 2 continue to apply to the second formulation of the problem. Furthermore, under Assumption 1, the consistency and asymptotic normality of the estimators without bound constraints, the consistency of the estimators with active bound constraints, and the results regarding the estimation of  $\sigma_p^2$  and  $\sigma_q^2$  hold.

*Proof.* See Appendix F □

Note that, like the unweighted estimators, the weighted estimators are consistent estimators of  $\mu_p$  and  $\mu_q$ . This means that although the unweighted and weighted estimators differ in finite samples, they are asymptotically equivalent. We use the weighted version of

our model in applications since, as was previously argued, it offers a better measure of the actual aggregate voter transition probabilities.

An important point related to this discussion is that, in most applications, we are interested in estimating *realized fractions*, as in Section 2, rather than the parameters we have considered in this section thus far. However, by comparing (1) and (3), wherein the first equation  $p_i$  and  $q_i$  represent fractions and in the second the unit level parameters, we can see that imposing the first moment condition (3) implies that we can treat  $\hat{p}_i$  and  $\hat{q}_i$  as estimates of the unit level fractions and  $\hat{p}$  and  $\hat{q}$  as estimates of the aggregate fractions across all units, which in Section 2 we also labeled  $\mu_p$  and  $\mu_q$ . This is another reason why the weighted versions of our estimators are preferable, since by weighting each unit appropriately, we can estimate the fractions in the aggregate voter transition matrix.

We now briefly show that our approach can also be extended to the RxC case, where, as discussed in Section 2, we divide voters in the first election into  $R$  groups and voters in the second election into  $C$  groups. In this case, we are interested in the fraction from each of these  $R$  groups that also belong to each of the  $C$  groups in the second election. If we define each group in each election by different candidates or competing options, this is equivalent to asking what fraction from each of the  $R$  groups in the first election vote for each of the  $C$  candidates or competing options in the second election. We can rewrite (4) for the RxC case in the following manner:

$$\begin{aligned}
& \min_{\{\hat{p}_{11}, \dots, \hat{p}_{RC}\} \in [0,1]^{RC}} \sum_{i=1}^N \left[ \sum_{r=1}^R \sum_{c=1}^C (\hat{p}_{rci} - \hat{p}_{rc})^2 \right] \\
& \text{s.t. } \{\hat{p}_{11i}, \dots, \hat{p}_{RCi}\} \in \underset{\{\hat{p}_{11i}, \dots, \hat{q}_{RCi}\} \in [0,1]^{RC}}{\text{argmin}} \left[ \sum_{r=1}^R \sum_{c=1}^C (\hat{p}_{rci} - \hat{p}_{rc})^2 \right], \quad i = 1, \dots, N \\
& \text{s.t. } \sum_{r=1}^R n_{ri} \hat{p}_{rci} = v_{ci}, \quad c = 1, \dots, C \\
& \sum_{c=1}^C \hat{p}_{rci} = 1, \quad r = 1, \dots, R
\end{aligned} \tag{8}$$

where  $n_{ri}$  is the number of votes for option  $r = 1, \dots, R$  in the first election in unit  $i$ ,  $v_{ci}$  is the number of votes for option  $c = 1, \dots, C$  in the second election in unit  $i$ , and  $\hat{p}_{rc}$  is our estimator for  $\mu_{rc}$ , the expected value of the probability that voters in group  $r$  in the first election vote for  $c$  in the second election across units. In this case, these probabilities satisfy  $\sum_{c=1}^C \mu_{rc} = 1$ , since each voter in each group  $r$  votes for one and only one option  $c$  in the second election. Using the same arguments as those in the proof of Proposition 1 in

Appendix A, we can turn (8) into a standard optimization problem in the following way:

$$\begin{aligned}
& \min_{\{\hat{p}_{11i}, \dots, \hat{p}_{RCi}\}_{i=1, \dots, N}} \sum_{i=1}^N \left[ \sum_{r=1}^R \sum_{c=1}^C (\hat{p}_{rci} - \frac{1}{N} \sum_{i=1}^N \hat{p}_{rci})^2 \right] \\
& \text{s.t.} \quad \sum_{c=1}^C \hat{p}_{rci} = 1, \quad i = 1, \dots, N, r = 1, \dots, R \\
& \quad \quad \sum_{r=1}^R n_{ri} \hat{p}_{rci} = v_{ci}, \quad i = 1, \dots, N, c = 1, \dots, C \\
& \quad \quad \hat{p}_{rci} \in [0, 1], \quad i = 1, \dots, N, r = 1, \dots, R, c = 1, \dots, C
\end{aligned} \tag{9}$$

By comparing (8) to (4) and (9) to (7), we can see that the RxC case is a straightforward extension of the 2x2 case. Further extending (8) and (9) to the case with weights is carried out in a manner analogous to Proposition 4, and we use the formulation of (9) with weights in one of the applications below.

We leave the extensions of all other properties listed in this section to the RxC case for future research, but we conjecture that, given the parallels between both problems, the estimators for the more general RxC case satisfy all of the properties that were listed here for the 2x2 case with proofs that follow along the same lines.

## 5. Simulation Study

This section presents simulations to evaluate the small sample behavior of our approach and compare it with that of the Bayesian hierarchical model in [King, Rosen, and Tanner \(1999\)](#). We present six simulations. In each case we consider  $N = 100$  units, and simulate data for each unit  $i$  independently as follows:

$$\begin{aligned}
p_i &\sim \text{Beta}(\alpha_p, \beta_p), \quad i = 1, \dots, N. \\
q_i &\sim \text{Beta}(\alpha_q, \beta_q), \quad i = 1, \dots, N. \\
n_i + m_i &\sim \text{Bin}(1000, 0.5), \quad i = 1, \dots, 100. \\
n_i / (n_i + m_i) &\sim \text{TN}(0.5, 0.12, 0.2, 0.8), \quad i = 1, \dots, N. \\
m_i / (n_i + m_i) &= 1 - n_i / (n_i + m_i), \quad i = 1, \dots, 100. \\
v_i &\sim \text{Bin}(n_i, p_i) + \text{Bin}(m_i, q_i), \quad i = 1, \dots, N.
\end{aligned}$$

where  $\text{TN}(\mu, \sigma, \text{LB}, \text{UB})$  denotes a truncated normal distribution with mean  $\mu$ , standard deviation  $\sigma$ , lower bound LB, and upper bound UB. We vary  $\alpha_p$  and  $\beta_p$  through the six cases and run 100 simulations for each parameter configuration. For each estimation, we use the weighted version of our model, with  $\psi_i = \frac{n_i + m_i}{\sum_{i=1}^N n_i + m_i}$ . Estimation for our model was carried out using the NLOPTR non-linear optimization library in R, while estimation

for the Bayesian hierarchical model in [King, Rosen, and Tanner \(1999\)](#) was carried out in Python using the PyEI library<sup>7</sup> ([Knudson, Schoenbach, and Becker, 2021](#)) setting the mean of the exponential hyperpriors to be 0.01<sup>8</sup>. We assessed convergence for the latter model with the  $\hat{R}$  and effective sample size (ESS) diagnostics, which are typical convergence diagnostics for MCMC sampling.

We present results for the average estimates for  $\mu_p$  and  $\mu_q$ , the root mean squared error (RMSE) for these estimates, the average root weighted mean squared error (RWMSE) for the unit-level estimates of  $p_i$  and  $q_i$ , and the average fraction of unit-level estimates  $\hat{p}_i$  and  $\hat{q}_i$  with active bound constraints. We compute the RWMSE rather than the RMSE for the unit-level estimates to account for the fact that some units have larger values of  $n_i + m_i$ . We also compute an unconstrained (UC) version of our model, in which we do not impose bound constraints, to provide some evidence related to [Corollary 1](#) and the result regarding both models' MSEs. Finally, we estimate  $\sigma_p^2$  and  $\sigma_q^2$  both with the estimators in [Corollary 2](#) and with the sample variance of our unit level estimates. [Table 3](#) compares our estimates of  $\mu_p$  and  $\mu_q$  and the unit-level parameters for each parameter configuration (simulation, for simplicity) to the unconstrained version of our model and the Bayesian hierarchical model. [Table 4](#) presents our mean estimates for  $\sigma_p$  and  $\sigma_q$  for each parameter configuration.

Both our model and the Bayesian hierarchical model successfully estimate  $\mu_p$  and  $\mu_q$ . Differences in the RMSE for these estimates between both models are generally minimal. We find that our model provides superior estimates for  $\mu_p$  and  $\mu_q$  when both true values are close to 0 and 1 (see, for example, [Simulation 3](#), [Simulation 4](#), and [Simulation 6](#) in [Table 3](#)). In contrast, the Bayesian hierarchical model has an equivalent or slightly superior performance when the true values are further from 0 and 1. For example, in [Simulation 1](#), [Simulation 2](#), and [Simulation 5](#) in [Table 3](#) our model's estimates are better for one parameter but worse for the other. In line with [Corollary 1](#), our model always performs better than its unconstrained version in this measure when there are some estimates with active bound constraints. In other cases, they are equivalent.

Our model tends to better estimate the unit-level values  $p_i$  and  $q_i$ . Its RWMSE is lower than the one resulting from the Bayesian hierarchical model in all but one of the simulations. In some simulations, the differences in this measure between both models can be substantial. For example, in [Simulation 6](#), the RWMSE for our model's estimates for  $p_i$  across units is approximately 45% lower than the corresponding RWMSE for the Bayesian

---

<sup>7</sup>Results were similar when we estimated the model in other ways.

<sup>8</sup>We found that this choice allowed the model to have a better performance than the value of 0.5 advocated for in [King, Rosen, and Tanner \(1999\)](#).

Table 3: Simulation Study - Comparison

|                           | Minimum Distance |         | Minimum Distance: UC |         | Hierarchical Model |         |
|---------------------------|------------------|---------|----------------------|---------|--------------------|---------|
|                           | $p$              | $q$     | $p$                  | $q$     | $p$                | $q$     |
| <b>Simulation 1</b>       |                  |         |                      |         |                    |         |
| True Values: $\mu$        | 0.70000          | 0.25000 | 0.70000              | 0.25000 | 0.70000            | 0.25000 |
| True Values: $\sigma$     | 0.03729          | 0.02417 | 0.03729              | 0.02417 | 0.03729            | 0.02417 |
| Mean Estimates: $\mu$     | 0.69878          | 0.25079 | 0.69878              | 0.25079 | 0.69957            | 0.25098 |
| RMSE: $\mu$               | 0.01295          | 0.01265 | 0.01295              | 0.01265 | 0.01227            | 0.01287 |
| RWMSE - Units             | 0.03390          | 0.03257 | 0.03390              | 0.03257 | 0.03899            | 0.03402 |
| Restricted Unit Estimates | 0.00000          | 0.00000 | 0.00000              | 0.00000 | -                  | -       |
| <b>Simulation 2</b>       |                  |         |                      |         |                    |         |
| True Values: $\mu$        | 0.80000          | 0.15000 | 0.80000              | 0.15000 | 0.80000            | 0.15000 |
| True Values: $\sigma$     | 0.03255          | 0.02058 | 0.03255              | 0.02058 | 0.03255            | 0.02058 |
| Mean Estimates: $\mu$     | 0.80018          | 0.15021 | 0.80018              | 0.15021 | 0.80113            | 0.14895 |
| RMSE: $\mu$               | 0.01165          | 0.01151 | 0.01165              | 0.01151 | 0.01197            | 0.01084 |
| Weighted RWMSE - Units    | 0.02963          | 0.02801 | 0.02963              | 0.02801 | 0.03522            | 0.02893 |
| Restricted Unit Estimates | 0.00000          | 0.00000 | 0.00000              | 0.00000 | -                  | -       |
| <b>Simulation 3</b>       |                  |         |                      |         |                    |         |
| True Values: $\mu$        | 0.91000          | 0.04000 | 0.91000              | 0.04000 | 0.91000            | 0.04000 |
| True Values: $\sigma$     | 0.02019          | 0.01237 | 0.02019              | 0.01237 | 0.02019            | 0.01237 |
| Mean Estimates: $\mu$     | 0.90809          | 0.04215 | 0.90951              | 0.04075 | 0.91687            | 0.03296 |
| RMSE: $\mu$               | 0.00674          | 0.00678 | 0.00749              | 0.00743 | 0.01001            | 0.00989 |
| RWMSE - Units             | 0.01861          | 0.01703 | 0.01905              | 0.01739 | 0.02670            | 0.01944 |
| Restricted Unit Estimates | 0.00000          | 0.00810 | 0.00000              | 0.00810 | -                  | -       |
| <b>Simulation 4</b>       |                  |         |                      |         |                    |         |
| True Values: $\mu$        | 0.95000          | 0.50000 | 0.95000              | 0.50000 | 0.95000            | 0.50000 |
| True Values $\sigma$      | 0.01147          | 0.02280 | 0.01147              | 0.02280 | 0.01147            | 0.02280 |
| Mean Estimates: $\mu$     | 0.94827          | 0.50234 | 0.94993              | 0.50065 | 0.95001            | 0.50028 |
| RMSE: $\mu$               | 0.00814          | 0.00926 | 0.00940              | 0.01028 | 0.01067            | 0.01075 |
| RWMSE - Units             | 0.02188          | 0.02548 | 0.02241              | 0.02606 | 0.02402            | 0.03199 |
| Restricted Unit Estimates | 0.00980          | 0.00000 | 0.00980              | 0.00000 | 0.00000            | -       |
| <b>Simulation 5</b>       |                  |         |                      |         |                    |         |
| True Values: $\mu$        | 0.60000          | 0.11111 | 0.60000              | 0.11111 | 0.60000            | 0.11111 |
| True Values: $\sigma$     | 0.02526          | 0.01047 | 0.02526              | 0.01047 | 0.02526            | 0.01047 |
| Mean Estimates: $\mu$     | 0.59935          | 0.11122 | 0.59935              | 0.11122 | 0.60181            | 0.10926 |
| RMSE: $\mu$               | 0.01094          | 0.01096 | 0.01094              | 0.01096 | 0.01157            | 0.01036 |
| RWMSE - Units             | 0.02788          | 0.02478 | 0.02788              | 0.02478 | 0.03426            | 0.02282 |
| Restricted Unit Estimates | 0.00000          | 0.00000 | 0.00000              | 0.00000 | -                  | -       |
| <b>Simulation 6</b>       |                  |         |                      |         |                    |         |
| True Values: $\mu$        | 0.95000          | 0.02500 | 0.95000              | 0.02500 | 0.95000            | 0.02500 |
| True Values: $\sigma$     | 0.01404          | 0.00581 | 0.01404              | 0.00581 | 0.01404            | 0.00581 |
| Mean Estimates: $\mu$     | 0.94820          | 0.02730 | 0.94961              | 0.02593 | 0.95627            | 0.01846 |
| RMSE: $\mu$ : $\mu$       | 0.00490          | 0.00476 | 0.00527              | 0.00511 | 0.00907            | 0.00861 |
| RWMSE - Units             | 0.01302          | 0.01186 | 0.01333              | 0.01209 | 0.02403            | 0.01585 |
| Restricted Unit Estimates | 0.00000          | 0.01370 | 0.00000              | 0.01370 | -                  | -       |

*Notes:* This table presents a comparison of the performance of our model, our model with no bound constraints, and the [King, Rosen, and Tanner \(1999\)](#) model across six different simulations. The rows contain the true values for  $\mu_p$  and  $\mu_q$ , the true values for  $\sigma_p$  and  $\sigma_q$ , the mean estimate for  $\mu_p$  and  $\mu_q$  across all simulations, the root mean squared error (RMSE) of  $\hat{p}$  and  $\hat{q}$  for each model, the mean root weighted mean squared error (RWMSE) across all simulations for the estimates of the unit level pairs  $(p_i, q_i)$ , and the average fraction of units where the bound constraints of  $\hat{p}_i$  and  $\hat{q}_i$  are active across all simulations. See Section 4 for details regarding the previous parameters and estimators.

Table 4: Simulation Study - Variance Estimates

|                            | $\sigma_p$ | $\sigma_q$ |
|----------------------------|------------|------------|
| <b>Simulation 1</b>        |            |            |
| True Values                | 0.03729    | 0.02417    |
| Estimates                  | 0.03643    | 0.02170    |
| Estimates: Sample Variance | 0.03011    | 0.02890    |
| <b>Simulation 2</b>        |            |            |
| True Values                | 0.03255    | 0.02058    |
| Estimates                  | 0.03092    | 0.01960    |
| Estimates: Sample Variance | 0.02585    | 0.02465    |
| <b>Simulation 3</b>        |            |            |
| True Values                | 0.02019    | 0.01237    |
| Estimates                  | 0.01841    | 0.01179    |
| Estimates: Sample Variance | 0.01637    | 0.01524    |
| <b>Simulation 4</b>        |            |            |
| True Values                | 0.01147    | 0.02280    |
| Estimates                  | 0.01165    | 0.02020    |
| Estimates: Sample Variance | 0.02012    | 0.02206    |
| <b>Simulation 5</b>        |            |            |
| True Values                | 0.02526    | 0.01047    |
| Estimates                  | 0.02414    | 0.00950    |
| Estimates: Sample Variance | 0.02331    | 0.02172    |
| <b>Simulation 6</b>        |            |            |
| True Values                | 0.01404    | 0.00581    |
| Estimates                  | 0.01308    | 0.00503    |
| Estimates: Sample Variance | 0.01160    | 0.01054    |

*Notes:* This table presents our model's estimates for  $\sigma_p$  and  $\sigma_q$  for six different simulations. For each simulation, rows contain the true values for  $\sigma_p$  and  $\sigma_q$ , the mean estimate for  $\sigma_p$  and  $\sigma_q$ , and the mean estimate for  $\sigma_p$  and  $\sigma_q$  using the sample variance of  $\hat{p}_i$  and  $\hat{q}_i$ . See Section 4 for details regarding the previous parameters and estimators.

model, while the RWMSE for  $q_i$  across units is approximately 25% lower. The RWMSE of our model is also always lower or equal to the one resulting from the unconstrained version of our model, further illustrating Corollary 1.

The estimates for the standard deviations are also relatively accurate. Our estimates for the variances computed as in Corollary 2 are very close to the true values in all cases, and there is no evidence of a systematic bias in any direction. On the other hand, the sample variance of the unit estimates  $\hat{p}_i$  and  $\hat{q}_i$  tend to overestimate the lowest standard deviation

and underestimate the largest standard deviation. The sample standard deviation of the unit estimates also tends to be very close to  $\frac{\sigma_p + \sigma_q}{2}$ . This implies that while the sample standard deviations may not be accurate estimates for  $\sigma_p$  and  $\sigma_q$  individually, they may provide us with an order of magnitude for the overall dispersion in  $p_i$  and  $q_i$  across units.

These results show that our model successfully estimates  $\mu_p$  and  $\mu_q$ , the main parameters we are interested in in most ecological inference applications. Importantly, there is no parameter configuration where the Bayesian hierarchical model is strictly better, while there are some where using our model yields substantial gains in accuracy. In particular, our model tends to be better with true values close to 0 and 1. The estimates for the standard deviations also perform well. If we want an alternative measure of the dispersion in  $p_i$  and  $q_i$  can use the sample variance of the unit estimates to obtain estimates for the true sample variance, keeping in mind that one of these estimates will tend to overestimate the true value and the other will tend to underestimate it.

## 6. Application: 2x2 Case

In this section, we apply the previous approach to two cases of 2x2 ecological inference with real data. In both cases, besides the aggregate data, the true values in the  $N$  tables such as Table 1 are known, so we can compare our estimates to the true values and compare the performance of our (weighted) model to the [King, Rosen, and Tanner \(1999\)](#) Bayesian hierarchical model. In all applications, we estimated both models in the same way as we did in Section 5. Note that these known quantities are realized fractions rather than underlying probabilities. In line with the discussion at the end of Section 4, in these applications we estimate these fractions rather than other parameters. To avoid introducing more notation, in what follows we refer to the true unit-level *fractions* by  $p_i$  and  $q_i$ , to the true aggregate fractions across all units by  $\mu_p$  and  $\mu_q$ , and to the sample variances of the unit-level fractions by  $\sigma_p^2$  and  $\sigma_q^2$ .

### 6.1. 2x2 Application: US

Our application to the US is slightly different from the examples of ecological inference that we have considered thus far in that no election is involved. However, it may be recast in the same framework. We divide the voting-age population (in what follows, voters) into two groups: voters of color and white voters. We then group voters into those who register to vote and those who do not register, and let  $p_i$  and  $q_i$  be the fraction of voters of color and white voters that register to vote. We are interested in  $\mu_p$ , the aggregate fraction of

voters of color who register to vote, and  $\mu_q$ , the aggregate fraction of white voters who register to vote. Therefore, we are estimating a voter transition matrix from white and colored voters into voter registration and non-registration.

The data is available at the county level (so that each unit corresponds to one county) for 268 counties in Florida, Louisiana, North Carolina, and South Carolina. It includes data on the total number of voters ( $n_i + m_i$ ), voters of color ( $n_i$ ), and the total number of registered voters ( $v_i$ ). We have access to the true values, so we also have data on the true fraction of white and colored voters who register to vote in each county. In particular, the true values for  $\mu_p$  and  $\mu_q$  are 0.58 and 0.82, respectively, so that on aggregate 58% of colored voters and 82% of white voters register to vote. Figure 2 plots the true pairs of fractions  $(p_i, q_i)$  for the 268 counties along with the  $(\mu_p, \mu_q)$  pair in red.

The figure shows that there is substantial heterogeneity in the true values. There are also many unit-level fractions that are 0 or 1, meaning there are counties with no voter registration by voters of color and counties where all white voters are registered to vote.

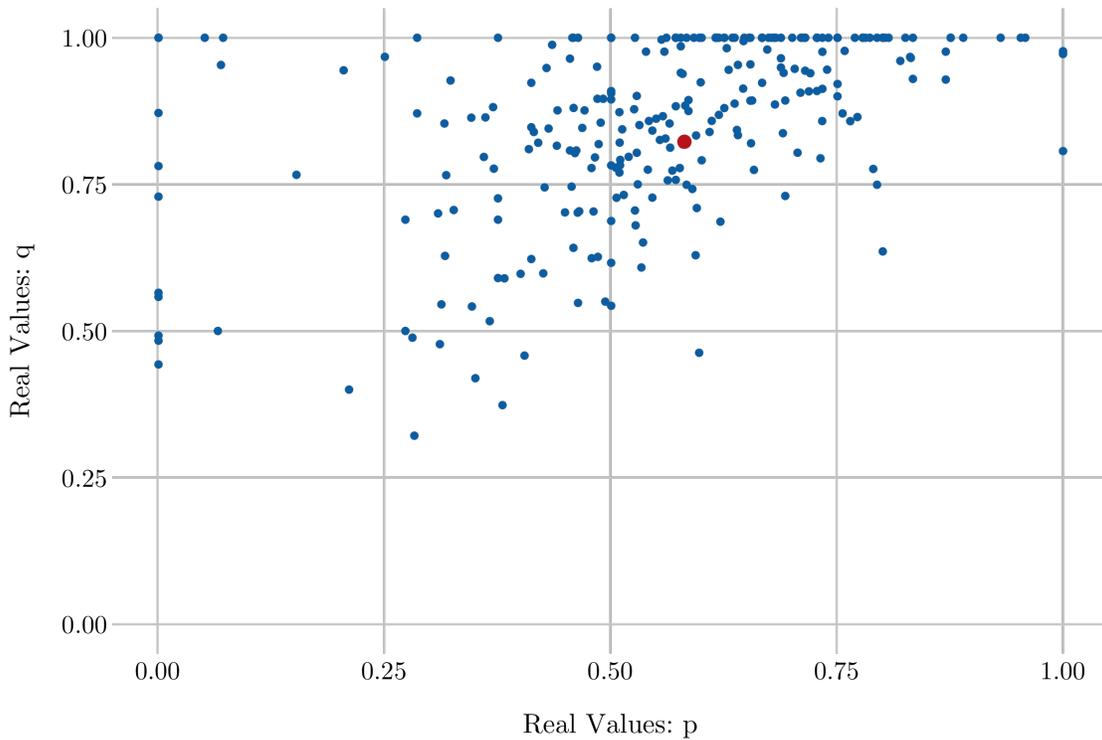


Figure 2: US 2x2 Application - True Values

*Notes:* This figure plots in blue the true values for  $p_i$ , the fraction of registered voters of citizens of color in unit  $i$ , and  $q_i$ , the fraction of registered white citizens in unit  $i$ , for 268 counties in four states in the US south. The red dot corresponds to the aggregate fractions  $\mu_p$  and  $\mu_q$  across all counties.

The true values for  $\mu_p$  and  $\mu_q$ , their estimates, and the RWMSE for the unit estimates are presented in Table 5 for our model and the [King, Rosen, and Tanner \(1999\)](#) hierarchical model. Table 6 presents results regarding the estimation of  $\sigma_p^2$  and  $\sigma_q^2$ .

Table 5: US 2x2 Application - Results

| Race                          | Total     | True Values          |                      | Voter Registration<br>Minimum Distance |                      | Hierarchical Model   |                      |
|-------------------------------|-----------|----------------------|----------------------|--|----------------------|----------------------|----------------------|
|                               |           | Yes                  | No                   | Yes                                    | No                   | Yes                  | No                   |
|                               |           | 6,748,677            | 2,095,023            | 6,748,677                              | 2,095,023            | 6,748,677            | 2,095,023            |
| POC                           | 1,961,476 | 0.58182<br>(0.20005) | 0.41818<br>(0.20005) | 0.58918<br>(0.03198)                   | 0.41082<br>(0.03198) | 0.58825<br>(0.04357) | 0.41175<br>(0.04357) |
| White                         | 6,882,224 | 0.82270<br>(0.15495) | 0.17730<br>(0.15495) | 0.81203<br>(0.01535)                   | 0.18797<br>(0.01535) | 0.81426<br>(0.01363) | 0.18574<br>(0.01363) |
| <b>RWMSE - Unit Estimates</b> |           | -                    |                      | $p_i$ : 0.12852                        | $q_i$ : 0.05316      | $p_i$ : 0.13929      | $q_i$ : 0.05137      |

*Notes:* This table presents the resulting estimates for the true values of the parameters using our model and the [King, Rosen, and Tanner \(1999\)](#) model. Sample standard deviations are in parentheses for the true values. Bootstrap standard errors using 10,000 samples are in parentheses for the Minimum Distance model. Standard deviations of the samples drawn out of the posterior distribution are in parentheses for the Bayesian hierarchical model.

Our model and the Bayesian hierarchical model provide relatively accurate estimates of  $\mu_p$  and  $\mu_q$ , with the Bayesian hierarchical model providing point estimates that are slightly closer to the true values than ours. However, they are not statistically different, and they are both within one standard deviation of the true values. The values for the weighted RWMSEs on the last row imply that our model and the Bayesian hierarchical model have a relatively similar performance. Specifically, our unit estimates in this application are superior overall but have a slightly larger error when estimating the fraction of white voters registering to vote.

Table 6: US 2x2 Application - Standard Deviation Estimates

|                            | $\sigma_p$ | $\sigma_q$ |
|----------------------------|------------|------------|
| True Values                | 0.20005    | 0.15495    |
| Estimates                  | 0.27925    | 0.13665    |
| Estimates: Sample Variance | 0.08639    | 0.14706    |

*Notes:* This table presents our estimates for  $\sigma_p$  and  $\sigma_q$ . These are the standard deviations for the fractions of voters of color and white voters who register to vote in four states in the US, respectively. In order, rows contain the true values for  $\sigma_p$  and  $\sigma_q$ , our estimates for  $\sigma_p$  and  $\sigma_q$ , and the estimate for  $\sigma_p$  and  $\sigma_q$  using the sample variance of  $\hat{p}_i$  and  $\hat{q}_i$ . See Section 4 for details regarding the previous parameters and estimators.

Our estimate of  $\sigma_q$  in Table 6 is close to the true value for  $\sigma_q$ , but our estimate for  $\sigma_p$  is far off the true value. The fact that many of the true values are either 0 or 1 and the high dispersion of these values shown in Figure 2 may make the estimation of these standard deviations difficult. However, they are closer overall to the true values than the estimates using the sample variance, consistent with the discussion in Section 4 and the results in Section 5.

Overall, our model estimates  $\mu_p$  and  $\mu_q$  accurately, which are the fractions composing the aggregate voter transition matrix and are, therefore, the values we are most interested in. It also partially succeeds in estimating  $\sigma_p$  and  $\sigma_q$ . In this particular application, however, there is no significant difference in accuracy between our model and the [King, Rosen, and Tanner \(1999\)](#) model.

## 6.2. 2x2 Application: Chile

We now illustrate our model with an application regarding voter turnout in Chile. In each election, we divide the voting-age population (or voters) into voters ( $n_i$ ) and non-voters ( $m_i$ ), as in the example we considered in Section 2. Let  $p_i$  and  $q_i$  be the fraction of first-election voters and non-voters that vote in the second election. We are interested in  $\mu_p$ , the aggregate fraction of first-round voters who vote in the runoff, and  $\mu_q$ , the aggregate fraction of first-round non-voters who vote in the runoff. This means we are estimating the aggregate voter transition matrix between voters and non-voters in the first round of the elections and voters and non-voters in the runoff.

The data consists of the voting-age population and the number of votes in the first round and runoff of the 2013 Chilean presidential election for each unit. As in the previous application, we have data on the true voter transition matrix. In this application, this means that for each unit we also have data on the number of first-round voters and non-voters who voted in the runoff. Each unit consists of one voting table of approximately 350 voters on average. For ease of exposition, we call these tables ballot boxes in the rest of this subsection.<sup>9</sup>

After eliminating ballot boxes with no votes either in the first round or the runoff, we are left with data on 41,094 ballot boxes. We use data on the 50 largest municipalities in the country, as measured by their number of voters. These municipalities encompass approximately 60% of all voters and contain approximately 20,000 ballot boxes. Figure 3

---

<sup>9</sup>In Chile, each citizen is assigned to a table with one ballot box within a voting locale. Each table groups approximately 350 voters, and votes are counted and logged for the ballot boxes in each table at the end of the election. Therefore, this consists of the lowest possible level of aggregation.

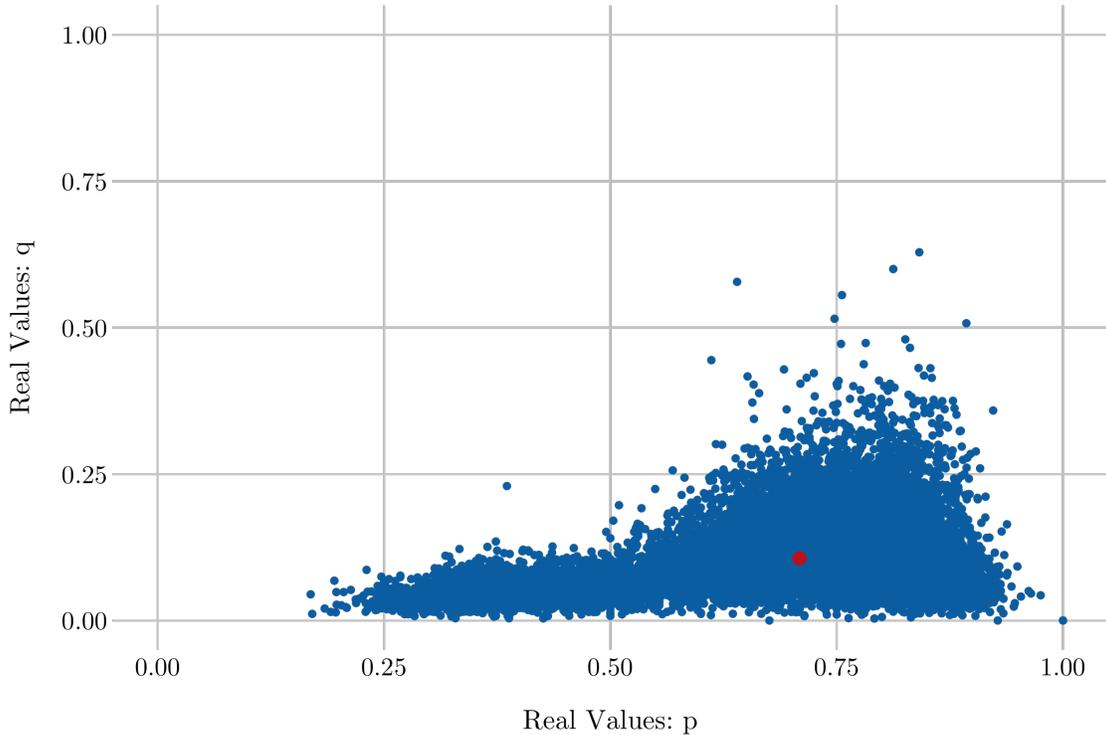


Figure 3: Chile 2x2 Application - True Values for All Ballot Boxes

*Notes:* This figure plots the true values for  $p_i$ , the fraction of first-round voters in the 2013 Chilean presidential election who voted in the runoff in unit  $i$ , and  $q_i$ , the fraction of first-round non-voters in the 2013 Chilean presidential election who vote in the runoff in unit  $i$ , for all ballot boxes in the 50 largest districts in the country.

plots the true pairs of fractions  $(p_i, q_i)$  for each ballot box in the final sample along with the true values for  $\mu_p$  and  $\mu_q$  in red. They are, respectively, 0.71 and 0.11, meaning that on aggregate 71% of first-round voters and 11% of first-round non-voters vote in the runoff.

Figure 3 shows that there are some outliers along the y-axis and that the variance along the x-axis is much larger than along the y-axis. This means that the decision to vote in the runoff for first-round voters is relatively heterogeneous across units. In contrast, the decision to vote for first-round non-voters is more homogenous and concentrated on lower values. Approximately 85% of the ballot boxes have a value of  $p_i$  of at least 0.6, which explains why the true values of  $\mu_p$  and  $\mu_q$  lie relatively far off to the right of the unit square despite the distribution of the true unit-level fractions.

We present three sets of estimates. First, we estimate the transition matrix using all ballot boxes in the sample together, meaning we estimate  $\mu_p$  and  $\mu_q$  using approximately 20,000 units. In this case, we present estimations for our model and Goodman regressions (see Section 2), given that such a large number of units makes the estimation of the

Bayesian hierarchical model in [King, Rosen, and Tanner \(1999\)](#) unviable. This is among the disadvantages of these approaches that were outlined in Section 3. For our second set of estimates, we estimate  $\mu_p$  and  $\mu_q$  separately for each of the 50 municipalities in the sample, using a number of ballot boxes ranging from 200 to 1000 per municipality. To estimate the true values for  $\mu_p$  and  $\mu_q$  for the whole sample, we then average these estimates over the 50 municipalities weighting each estimate proportionally to the number of voters in each. We present estimations for our model and the [King, Rosen, and Tanner \(1999\)](#) model. In our final set of estimates, we show how our model may be extended to deal with multiple clusters in the true values of  $p_i$  and  $q_i$ .

### All Ballot Boxes

We first present results for our model and for Goodman regressions estimated using all 20,000 units in the sample together. Table 7 shows the estimation results and the RWMSE for the unit estimates, while Table 8 presents results regarding the estimation of  $\sigma_p^2$  and  $\sigma_q^2$ .

Table 7: Chile 2x2 Application - Results by Ballot Boxes

|                               |              | Runoff               |                      |                      |                      |                       |                      |
|-------------------------------|--------------|----------------------|----------------------|----------------------|----------------------|-----------------------|----------------------|
|                               |              | True Values          |                      | Minimum Distance     |                      | Goodman Regression    |                      |
| First Round                   |              | Votes                | Abstains             | Votes                | Abstains             | Votes                 | Abstains             |
|                               | <b>Total</b> | 3,138,010            | 4,700,772            | 3,138,010            | 4,700,772            | 3,138,010             | 4,700,772            |
| Votes                         | 3,778,113    | 0.70894<br>(0.13436) | 0.29106<br>(0.13436) | 0.71019<br>(0.00111) | 0.28981<br>(0.00111) | 0.86907<br>(0.00391)  | 0.13093<br>(0.00391) |
| Abstains                      | 4,060,669    | 0.10611<br>(0.05436) | 0.89389<br>(0.05436) | 0.10726<br>(0.00075) | 0.89274<br>(0.00075) | -0.05700<br>(0.00350) | 1.05700<br>(0.00350) |
| <b>RWMSE - Unit Estimates</b> |              | -                    |                      | $p_i$ : 0.07093      | $q_i$ : 0.05264      | $p_i$ : 0.20966       | $q_i$ : 0.1717       |

*Notes:* This table presents the resulting estimates for the true values for  $\mu_p$  and  $\mu_q$  using our model and [Goodman \(1953\)](#) regressions. Sample standard deviations are in parentheses for the true values. Bootstrap standard errors using 10,000 samples are in parentheses for the Minimum Distance model. Robust standard errors are in parentheses for the Goodman regressions.

As expected given the large number of units, estimates from our model are quite accurate for both  $\mu_p$  and  $\mu_q$ . In particular, our estimates are within two standard deviations from the true values. This is not the case when estimating  $\mu_p$  and  $\mu_q$  with Goodman regressions. The estimate for  $\mu_p$  is much larger, while the resulting estimate for  $\mu_q$  is negative. None of these estimates are within two standard deviations of the true values. This highlights that Goodman regressions are often inadequate for ecological inference, given that it returns estimates outside the  $[0, 1]$  interval for the relevant parameters despite having been

Table 8: Chile 2x2 Application - Standard Deviation Estimates by Ballot Boxes

|                            | $\sigma_p$ | $\sigma_q$ |
|----------------------------|------------|------------|
| True Values                | 0.13436    | 0.05436    |
| Estimates                  | 0.05379    | 0.07146    |
| Estimates: Sample Variance | 0.08247    | 0.05424    |

*Notes:* This table presents our estimates for  $\sigma_p$  and  $\sigma_q$ . These are the standard deviations for the fractions of first-round voters and non-voters who vote in the runoff to the 2013 Chilean presidential election, respectively. In order, rows contain the true values for  $\sigma_p$  and  $\sigma_q$ , our estimates for  $\sigma_p$  and  $\sigma_q$ , and the estimate for  $\sigma_p$  and  $\sigma_q$  using the sample variance of  $\hat{p}_i$  and  $\hat{q}_i$ . See Section 4 for details regarding the previous parameters and estimators.

estimated with approximately 20,000 units and a relatively low level of aggregation. Its large RWMSE for its unit estimates results from the assumption of uniform fractions  $p_i$  and  $q_i$  across units.

Our estimates for  $\sigma_p$  and  $\sigma_q$ , presented in Table 8, are relatively accurate for  $\sigma_q$  but not for  $\sigma_p$ . This is likely a consequence of the odd distribution of the true values in Figure 3, which might make accurately estimating these parameters difficult. In this case, the sample standard deviations of our unit estimates provide more accurate estimates for the sample standard deviations of the true unit-level fractions.

As with the previous application, the estimation of  $\mu_p$  and  $\mu_q$  is carried out successfully. Furthermore, we conducted our estimation using a number of units that did not allow for the use of more computationally intensive models, highlighting one of the advantages of our approach over others in applications with a relatively low level of aggregation such as this.

### By District

We now present the results for our second set of estimates. In this case, we first estimate  $\mu_p$  and  $\mu_q$  within each municipality individually. We then compute an estimate of these fractions for the whole sample by taking a weighted average of the previous estimates with weights proportional to the number of voters in each municipality. Since we are estimating  $\mu_p$  and  $\mu_q$  within municipalities as our first step, in Figure 4 we plot the  $(p_i, q_i)$  pairs for four of the largest municipalities as colored points along with the  $(p_i, q_i)$  pairs for the whole sample in the background. The same general pattern in Figure 3 holds for all four municipalities. However, in some, the pairs  $(p_i, q_i)$  are more clearly separated into two groups. For instance, in the panel on the bottom left, for La Florida, they form two

clusters. This suggests that the estimation of  $\mu_p$  and  $\mu_q$  within each municipality may not be as precise as the estimates presented in the previous section for all ballot boxes since we are estimating bunching all ballot boxes together rather than using the fact that there are two somewhat clearly distinguished clusters. This was not the case in the previous set of estimates, since using all ballot boxes together blurs that there seem to be two clusters in some districts, leading to a single (dispersed) cluster in the data. However, the case of the bottom-left panel is not true in all districts. As a result, the aggregate estimates resulting from a weighted average of  $\hat{p}$  and  $\hat{q}$  across the fifty municipalities still provide accurate estimates of the true values for the whole sample, as shown below.

Results are presented in Table 9. In this case, we do not present results regarding the estimation of  $\sigma_p^2$  and  $\sigma_q^2$ . For our model, the point estimates when using this estimation strategy are not statistically different from the point estimates in the previous case. The standard errors of our estimates are slightly smaller in this application, and the point estimates remain extremely close to the true values. Table 9 also shows that the Bayesian hierarchical model leads to estimates far off the true values. This could be due to our model being more robust to the presence of multiple clusters in some municipalities (such as La Florida in Figure 4), in the sense that the precision of our model’s estimates is less affected by the presence of multiple clusters. Consequently, the RWMSE for both unit estimates is much lower for our model.

Table 9: Chile 2x2 Application - Results by Districts

|                               |              | Runoff               |                      |                      |                      |                      |                      |
|-------------------------------|--------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|                               |              | True Values          |                      | Minimum Distance     |                      | Hierarchical Model   |                      |
| First Round                   |              | Votes                | Abstains             | Votes                | Abstains             | Votes                | Abstains             |
|                               | <b>Total</b> | 3,138,010            | 4,700,772            | 3,138,010            | 4,700,772            | 3,138,010            | 4,700,772            |
| Votes                         | 3,778,113    | 0.70894<br>(0.13436) | 0.29106<br>(0.13436) | 0.71069<br>(0.00108) | 0.28931<br>(0.00108) | 0.81007<br>(0.00069) | 0.18993<br>(0.00069) |
| Abstains                      | 4,060,669    | 0.10611<br>(0.05436) | 0.89389<br>(0.05436) | 0.10408<br>(0.00075) | 0.89592<br>(0.00075) | 0.01013<br>(0.00058) | 0.98987<br>(0.00058) |
| <b>RWMSE - Unit Estimates</b> |              | -                    |                      | $p_i$ : 0.07620      | $q_i$ : 0.06272      | $p_i$ : 0.11588      | $q_i$ : 0.10767      |

*Notes:* This table presents the resulting estimates for the true values of  $\mu_p$  and  $\mu_q$  for the whole sample using our model and the King, Rosen, and Tanner (1999) model. Sample standard deviations are in parentheses for the true values. Bootstrap standard errors using 10,000 samples are in parentheses for the Minimum Distance model. Standard deviations of the samples drawn out of the posterior distribution are in parentheses for the Bayesian hierarchical model.

Overall, our model performs much better than the Bayesian hierarchical model in King, Rosen, and Tanner (1999). Results are also virtually unchanged compared to our estimates in the previous subsection, where we used all ballot boxes in one go (Table

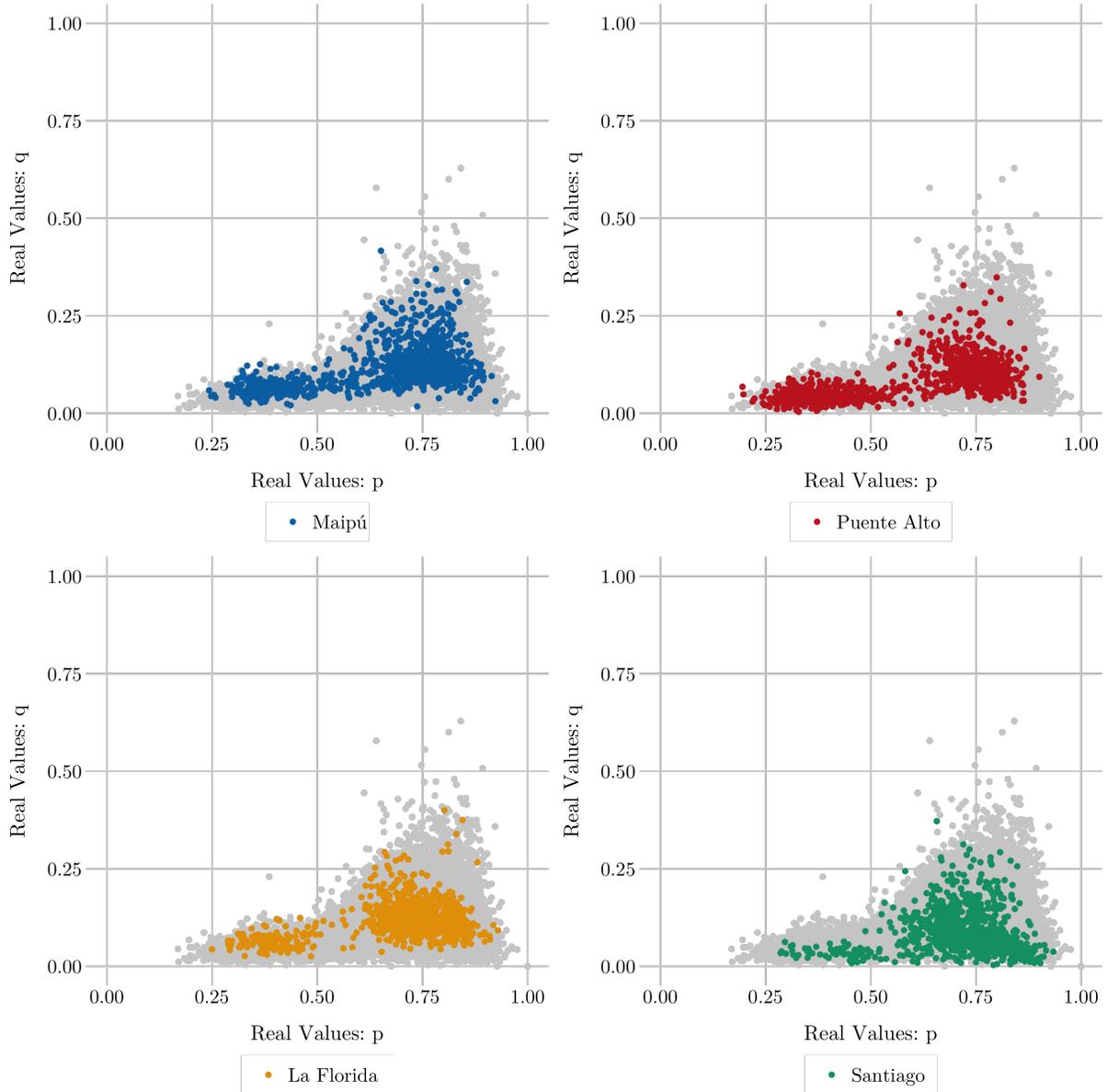


Figure 4: Chile 2x2 Application - True Values for Four Selected Municipalities

*Notes:* This figure plots the true values for  $p_i$ , the fraction of first-round voters in the 2013 Chilean presidential election who vote in the runoff in unit  $i$ , and  $q_i$ , the fraction of first-round non-voters in the 2013 Chilean presidential election who vote in the runoff in unit  $i$ , for all ballot boxes in four of the fifty largest municipalities in the country. The grey points are the true values for the whole sample, while the colored points in each panel are the true values for each specific municipality.

7). The apparent two clusters in some municipalities in Figure 4 may lead to imprecise estimates within districts, but they are not significant enough to affect the precision of the aggregate estimates. However, this may not always be the case, and different approaches are required to deal correctly with ecological inference in the presence of more than one

cluster. We explore one such possibility in our final set of results for this application.

## Multiple Clusters

As was previously discussed, Figure 4 shows that some municipalities have two clusters of true values of  $p_i$  and  $q_i$ , while our model has thus far been estimated without taking this feature of the data into account. This section shows that our model can be extended to deal with this case in a straightforward manner.

We proceed in the following way. First, we estimate our model using all available ballot boxes. This will result in estimates for the true fractions for each unit, i.e., for each ballot box. Then, we use these estimates to identify two clusters of ballot boxes using k-means clustering and estimate our model within each cluster. This, in turn, will result in estimates for  $\mu_p$  and  $\mu_q$  within each cluster. To obtain estimates for  $\mu_p$  and  $\mu_q$  for the municipality as a whole, we take a weighted average of our estimates for  $\mu_p$  and  $\mu_q$  in each of the clusters weighting each estimate proportionally to the number of voters in each cluster. The crucial requirement for this procedure to work is that we correctly identify the two clusters in the true values of  $p_i$  and  $q_i$  using our estimates for these fractions rather than their true values.

We illustrate the previous strategy using the example of La Florida (see Figure 4) and estimate  $\mu_p$  and  $\mu_q$  in three different ways. First, we use all available ballot boxes and ignore the issue of multiple clusters. Second, we use the previously outlined procedure. Third, we use the previously outlined procedure but identify the two clusters using the true values for  $p_i$  and  $q_i$  rather than our estimates. We provide these last sets of estimates to compare the results of our procedure with those of identifying each cluster using the true values. Results are summarized in Table 10.

Table 10: Chile 2x2 Application - Clusters (La Florida)

|                               | No Clusters     |                 | Estimated Clusters |                 | True Clusters   |                 |
|-------------------------------|-----------------|-----------------|--------------------|-----------------|-----------------|-----------------|
|                               | $\mu_p$         | $\mu_q$         | $\mu_p$            | $\mu_q$         | $\mu_p$         | $\mu_q$         |
| True Values                   | 0.69889         | 0.11862         | 0.69889            | 0.11862         | 0.69889         | 0.11862         |
| Estimates                     | 0.68608         | 0.13731         | 0.69778            | 0.11842         | 0.69766         | 0.11818         |
| Difference                    | 0.01281         | 0.01869         | 0.00111            | 0.00020         | 0.00123         | 0.00044         |
| <b>RWMSE - Unit Estimates</b> | $p_i$ : 0.06730 | $q_i$ : 0.06033 | $p_i$ : 0.04343    | $q_i$ : 0.04703 | $p_i$ : 0.04203 | $q_i$ : 0.04667 |
| <b>SD - Unit Estimates</b>    | $p_i$ : 0.08366 | $q_i$ : 0.06213 | $p_i$ : 0.03196    | $q_i$ : 0.02528 | $p_i$ : 0.03233 | $q_i$ : 0.02565 |

*Notes:* This table presents a comparison of the estimates when using clusters for the case of La Florida. In order, the rows contain the true values for  $\mu_p$  and  $\mu_q$ , the estimates from the model in the column, the absolute value of the difference between the estimate and the real value, the population-weighted RWMSE for the unit estimates, and the standard deviation of the unit estimates.

First, note that estimating separately by clusters before computing the aggregate estimates leads to significant gains in accuracy whether these clusters are identified using the estimated values for the unit-level fractions or their true values. The point estimates for  $\mu_p$  and  $\mu_q$  are much closer to the true values, and the RWMSE for the unit-level estimates drops by between 22% and 35% when clustering using either the estimated or true values. This is to be expected, given that our model may not return accurate estimates (despite them being consistent) in the case of multiple clusters. Second, the results of the estimated and real clusters are very similar. The point estimates are slightly better for the case with estimated clusters, but the RWMSEs for the unit estimates of  $p_i$  and  $q_i$  within La Florida are lower when using the true clusters. This implies that the clusters identified using the estimated values are almost exactly the same as those identified using the true values, which are unobserved in a typical ecological inference application, and constitutes some evidence that our approach may work in the general case in correctly identifying clusters of units in the data. We also present the standard deviation of the unit-level estimates to show the gain in precision that may be achieved when estimating separately for the two clusters. The standard deviation in each case with clusters drops by more than half when compared to the “No Clusters” case, which illustrates the gains in efficiency that are achievable by this approach. Furthermore, the standard deviations with the estimated clusters are very similar to those with the true clusters, further showing that our model is able to correctly identify both clusters.

Overall, this is a promising method for dealing with the issue of multiple clusters. The case of multiple clusters is empirically relevant, and it may arise in several applications such as the above. Furthermore, these applications are the ones where our approach may offer further advantages to other ecological inference methods. The estimation of two or more clusters separately may not be viable with high computational costs, and different approaches to deal with multiple clusters in other ecological inference models either rely on being robust to the presence of multiple clusters or on extensions that compound on the issue of their computational intensity. The approach presented here is free of both of these issues. However, we will not focus on the case of multiple clusters in this paper but explore this extension of our model in more detail in future work.

## 7. Application: RxC Case

We now use our model to estimate voter transitions between the runoff of the 2021 Chilean presidential elections and the 2022 Chilean Constitutional Plebiscite. We first

offer some context. In the 2021 presidential election runoff, the competing candidates were Gabriel Boric, a left-wing candidate representing a relatively new political coalition consisting of several non-traditional left-wing parties, and José Antonio Kast, a right-wing candidate likewise representing a new political movement outside of traditional right-wing parties. Both candidates were on the extremes of the political spectrum. The candidates representing conventional center-left and center-right coalitions came in fourth and fifth, respectively, in the first round of the elections<sup>10</sup>. In the 2022 Constitutional Plebiscite, the two competing options were to approve or reject the draft for a new constitution presented by the Constitutional Convention after a year of work. Leading to the elections, several controversies involving the elected representatives to the Convention and some of the articles that were included in the draft, such as the one declaring Chile a plurinational state, led to the most likely outcome being its rejection, despite the overwhelming result in the 2020 Constitutional Plebiscite two years prior that led to the establishment of the Constitutional Convention.

To estimate voter transitions between both elections, we group the voting-age population (in what follows, voters) in each election into three categories, making this an RxC ecological inference problem in the same fashion as that described at the end of Section 2. In the presidential election runoff, voters are split between Boric voters, Kast voters, and a group called Nulls, Blanks, and Abstentions (N-B-A), in which we group null votes, blank votes, and those voters that chose not to vote. In the 2022 plebiscite, voters are likewise split among Approve, Reject, and N-B-A voters. As in the 2x2 applications, we estimate the realized fractions for these voter transitions rather than underlying probabilities. In particular, we are interested in the fractions of Boric, Kast, and N-B-A voters that voted for each of the three alternatives in the 2022 plebiscite, both at the aggregate level and at the level of each unit.

## 7.1. Data

We use data from the Chilean Electoral Service (Servel) at the municipal level<sup>11</sup> for each election, which includes data on the number of voters for each option and the number of

---

<sup>10</sup>The candidate that came in third place, representing the populist *Partido de la Gente*, ran his whole campaign from the US due to a standing alimony payment issue in Chile with his ex-wife.

<sup>11</sup>Like in the previous application for Chile, the lowest level of aggregation available is at the ballot box level. However, the assignment to ballot boxes changed significantly between elections, so we cannot use data at this level. Therefore, we group the ballot box data at the level of each municipality, which is the next possible lowest level of aggregation. Doing this significantly reduces the chances of this change being an issue for estimation.

voters in each election<sup>12</sup>. Table 11 presents the election results.

Table 11: 2021 Presidential Election Runoff and 2022 Plebiscite Results

|                                  | Number     | Percentage |
|----------------------------------|------------|------------|
| <b>Presidential Runoff</b>       |            |            |
| Boric                            | 4,596,579  | 30.73%     |
| Kast                             | 3,640,606  | 24.34%     |
| N-B-A                            | 6,722,771  | 44.94%     |
| Voting Age Population            | 14,959,956 | 100%       |
| <b>Constitutional Plebiscite</b> |            |            |
| Approve                          | 4,826,237  | 32.01%     |
| Reject                           | 7,862,072  | 52.15%     |
| N-B-A                            | 2,387,313  | 15.84%     |
| Voting Age Population            | 15,075,622 | 100%       |

*Notes:* This table contains the results from the 2021 Presidential Runoff between Gabriel Boric and José Antonio Kast and for the 2022 Constitutional Plebiscite. In both elections, the “N-B-A” category groups null votes, blank votes, and voters who did not vote.

Note that there is a drastic decrease in the number of N-B-A voters between both elections. This is due to the presidential election runoff being conducted under a voluntary voting policy and a compulsory voting policy being implemented for the constitutional plebiscite. Since voter registration in Chile is automatic, this meant that all citizens of voting age had to vote or risk the possibility of a fine. There was also a significant increase in the number of Reject voters compared to the number of Kast voters. These results suggest that most Boric voters opted to approve the draft, while most Kast and N-B-A voters opted to reject it.

One issue for the estimation of our model is that, as shown in Table 11, there was an increase in the number of voters between both elections. Given that the time gap between both elections is less than a year, the increase is marginal compared to the total number of voters. However, as explained in Section 4, estimating our model requires the same number of voters in both elections. To deal with this issue, we artificially modify the total number of votes for each option in the runoff of the presidential election while keeping the results in percentage terms unchanged. This should not significantly affect our results since (i) the increase in the number of voters is only a small percentage of the total number of votes

<sup>12</sup>Data from the presidential runoff corresponds to the official results, while data from the constitutional plebiscite corresponds to provisional data from Servel. Provisional data leads only to insignificant differences in the total number of votes for each option, which have a negligible effect on the results.

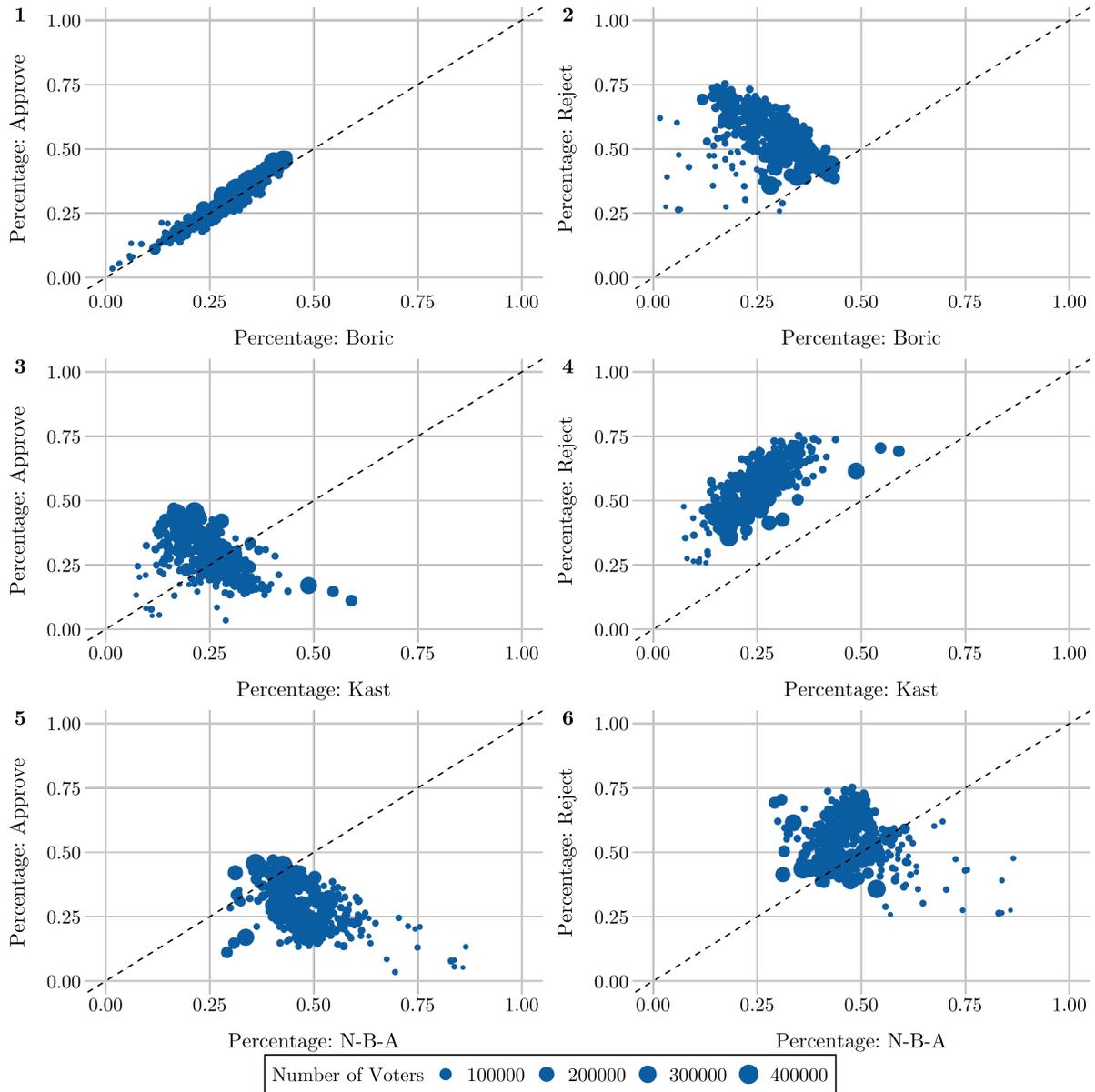


Figure 5: Presidential Election Runoff vs Constitutional Plebiscite Percentage Vote

*Notes:* Each panel plots the percentage obtained by each candidate in each unit in the runoff to the presidential election against the percentage obtained by each of the two main options in the constitutional plebiscite. The size of each point in the scatter plot represents the total number of voters in the runoff in each unit. Official and provisional results for the runoff and the constitutional plebiscite, respectively, were obtained at the ballot-box level from the Chilean Electoral Service.

and (ii) identification of the transition matrices for each unit and at the aggregate level depends mainly on the percentage of votes obtained by each option in either election<sup>13</sup>.

<sup>13</sup>An alternative to the previous way of dealing with the change in the electorate is to incorporate the change by modifying the number of people in the N-B-A category in the first election. For example, if the

Figure 5 shows the relation between the percentage of total votes obtained by each option in the runoff in each unit with the two main options in the constitutional plebiscite, which provides further evidence regarding voter transitions between both elections along with some caveats. First, while the percentage of Boric and Approve voters and Kast and Reject voters have a strong linear relationship (see panels 1 and 4), such a linear relation is absent in the other cases. For instance, there are several outliers in the relationship between the percentage vote for Boric and the percentage vote to Reject. Second, the percentage vote for the Reject option is larger than the percentage vote for Kast in every municipality, but the same does not hold for the support for the approval of the draft and the support for Boric. The outliers in each of the panels are units with a small number of voters and three units where the support for Kast and the option to reject was particularly strong (Las Condes, Vitacura, and Lo Barnechea). These three units consist of the three wealthiest municipalities in Chile. In our estimation, we use only the 150 municipalities with the largest number of voters, excluding Las Condes, Vitacura, and Lo Barnechea, as a way to eliminate most of these outliers<sup>14</sup>. The sample encompasses slightly under 85% of the total number of voters.

## 7.2. Results

We present estimates for the weighted version of our model for the RxC case. In particular, we use the weighted version of (9), weighting units proportionally to their number of voters and computing our estimates using the NLOPT non-linear optimization library in R. We also estimate the Rosen et al. (2001) RxC Bayesian hierarchical model using the PyEI library in Python<sup>15</sup> (Knudson, Schoenbach, and Becker, 2021) and assess convergence in the same way as we did in Section 5. Table 12 presents the results for both models. Since the true values are unknown, we cannot present the true transition matrix for this application.

Our results are consistent with the previous discussion. First, approximately 92% of Boric and 94% of Kast voters voted to approve and reject the constitutional draft,

---

number of N-B-A votes in the first election was two, but there was an increase in the electorate between both elections of two voters, we change the number of N-B-A votes in the first election to 4. Since voters who were incorporated between elections did not vote in the first election, this fits the definition of the N-B-A category given in the main text. Results (not shown) when using this way of dealing with the change in the electorate are very similar for both models and do not change the analysis.

<sup>14</sup>Including or excluding Las Condes, Vitacura, and Lo Barnechea has virtually no effects on the results of either of the models that we present in this section.

<sup>15</sup>Results for the Rosen et al. (2001) model were similar when using the eiPack library (Lau, Moore, and Kellermann, 2007) in R.

Table 12: Chile 3x3 Application - Results and Comparison

| <b>Presidential Runoff</b> |           | <b>Constitutional Plebiscite</b> |                      |                      |                                    |                      |                      |
|----------------------------|-----------|----------------------------------|----------------------|----------------------|------------------------------------|----------------------|----------------------|
|                            |           | <b>Minimum Distance</b>          |                      |                      | <b>Bayesian Hierarchical Model</b> |                      |                      |
|                            |           | Approve                          | Reject               | N-B-A                | Approve                            | Reject               | N-B-A                |
| <b>Total</b>               | 4,826,237 | 7,862,072                        | 2,387,313            | 4,826,237            | 7,862,072                          | 2,387,313            |                      |
| Boric                      | 4,596,579 | 0.92410<br>(0.01433)             | 0.05360<br>(0.01074) | 0.02230<br>(0.00447) | 0.95667<br>(0.00768)               | 0.03229<br>(0.00748) | 0.01104<br>(0.00567) |
| Kast                       | 3,640,606 | 0.01792<br>(0.00565)             | 0.94277<br>(0.04600) | 0.03931<br>(0.04418) | 0.02982<br>(0.01596)               | 0.88533<br>(0.03273) | 0.08485<br>(0.03080) |
| N-B-A                      | 8,329,332 | 0.08269<br>(0.00998)             | 0.60461<br>(0.03218) | 0.31270<br>(0.02674) | 0.05737<br>(0.00898)               | 0.64676<br>(0.01694) | 0.29587<br>(0.01595) |

*Notes:* This table presents the resulting estimates for the voter transitions from the 2021 presidential runoff to the 2022 Constitutional Plebiscite using our model and the [Rosen et al. \(2001\)](#) model. Bootstrap standard errors using 1000 samples are in parentheses for the Minimum Distance model. Standard deviations of the samples drawn out of the posterior distribution are in parentheses for the Bayesian hierarchical model.

respectively. This is consistent with panels 1 and 4 of Figure 5. However, the fraction of Boric voters who voted to reject is higher than that of Kast voters who voted to approve, which suggests that the Approve option was slightly less successful in gaining the vote of sympathizing groups. Furthermore, of the large number of voters that voted either to approve or reject the draft in the constitutional plebiscite but belong to the N-B-A group in the presidential runoff, approximately 88% voted to reject and only 12% to approve the constitutional draft. This is important since it is likely that these voters were brought on by the compulsory voting policy implemented for the constitutional plebiscite.

Table 12 also shows that our results are similar to those of the [Rosen et al. \(2001\)](#) model, although there are some differences in the point estimates, particularly in the Kast to N-B-A transition and in the N-B-A to Approve and N-B-A to Reject transitions. For the Kast to N-B-A transition, the point estimates for our model fall within the 95% credible intervals for the corresponding Bayesian estimates, meaning there is no significant difference between these estimates. This is not the case for the other two transitions, but these differences do not alter the interpretation of the results. In both cases, most runoff N-B-A voters rejected the constitutional draft. Furthermore, since most of the estimates for these transitions are relatively close to 0 and 1 and given the results of Section 5, there is reason to believe that our model may be more accurate in this particular application.

Figure 6 presents the unit-level estimates from our model for the transitions from each option in the runoff into the two main options in the constitutional plebiscite. For each

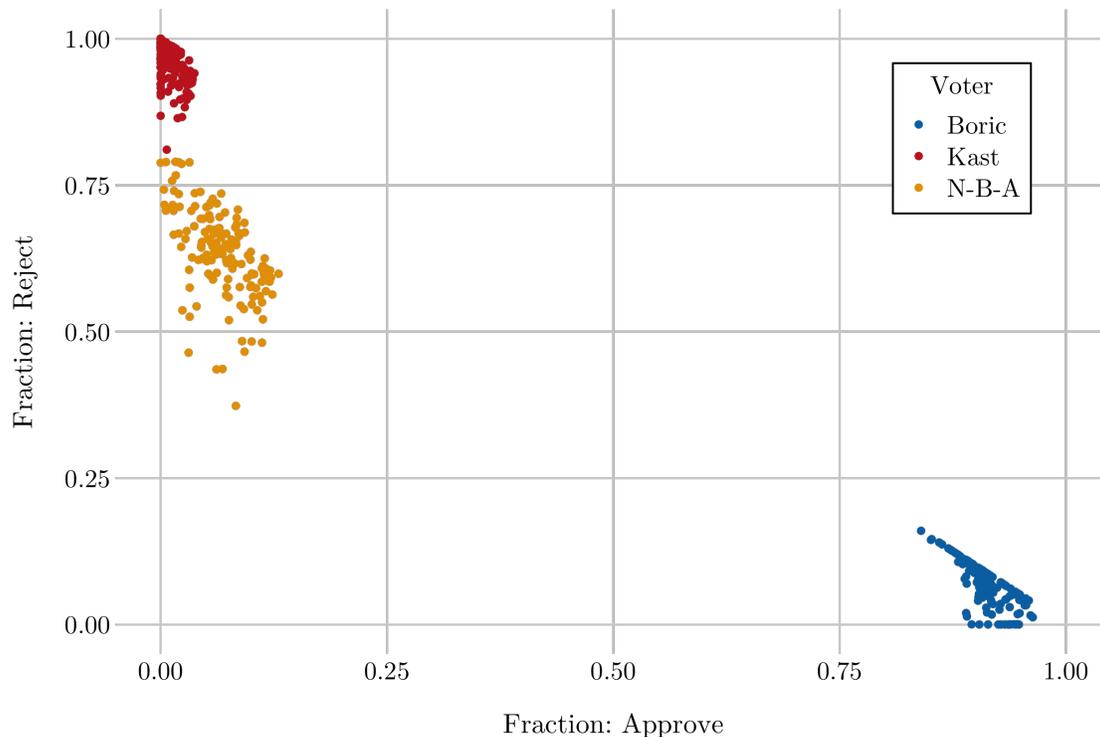


Figure 6: Chile 3x3 Application - Unit Level Estimates for the Minimum Distance Model

*Notes:* This figure plots the unit-level estimates for the voter transitions between the runoff of the 2021 presidential election and the 2022 constitutional plebiscite resulting from our model. Each point represents one group of voters in the runoff for one unit. Its position along the x-axis indicates the estimated fraction of voters in a given group that voted to approve the draft, and each point's position along the y-axis indicates the estimated fraction of voters in a given group that voted to reject the draft.

unit, each point represents one group of voters in the runoff. Its position along the x-axis indicates the estimated fraction of voters from that group that voted to approve, while its position along the y-axis indicates the estimated fraction of voters from that group that voted to reject.

Several interesting results follow from this figure. First, Boric voters, predictably, have a strong tendency to vote to approve the draft in all units. However, there is also some heterogeneity in their choices. In particular, the unit-level estimates for the fraction of Boric voters who voted to reject range from 0% to approximately 16%. Kast voters, on the other hand, are more homogenous. Our estimates for the fraction of Kast voters who voted to approve the draft range from 0% to just 4% across all units, and the median is approximately 1%. This further illustrates the point made from Table 12 that Boric voters were more likely to reject the draft than Kast voters were to approve it.

N-B-A voters from the runoff display two patterns. First, the estimates for the fraction

of N-B-A voters that voted to approve range from 0% to approximately 13%, meaning that the vote to approve among this group of voters is strikingly low in all units. The range of these estimates across all units is actually lower than that for the fraction of Boric voters who voted to reject the draft. On the other hand, the estimates for the N-B-A to Reject transition range from approximately 37% up to 79%, with the 25th percentile for these estimates at 59% and the median at 64%. Along with the estimates for the N-B-A to Approve transition, these results imply that runoff N-B-A voters voted overwhelmingly for the Reject option in most units. This also suggests that the primary source of heterogeneity in the voting patterns of N-B-A voters results from the decision to either not vote despite the compulsory voting policy<sup>16</sup> or voting to reject, while the decision to approve is more stable at low values across all units.

These results and those in Table 12 have significant implications regarding the analysis of these results and the impact of the compulsory voting policy implemented for the 2022 plebiscite. First, the discussion around Figure 6 suggests that voters who had not voted in the presidential election runoff and were thus likely brought on by the compulsory voting policy in place in the 2022 plebiscite are either (i) relatively more right-leaning than the rest of the population or (ii) “reactionary” voters who vote for the option that punishes the incumbent, which in these elections meant voting to reject the draft for the new constitution. Even though we cannot distinguish between these hypotheses based on our results alone, this is an important contribution to further understanding the results of these elections, providing to our knowledge the first evidence of these tendencies among new voters<sup>17</sup>.

Second, since most new voters (the N-B-A group in the presidential election runoff) voted to Reject, a plausible scenario is that the Reject option would have lost much of its support without a compulsory voting policy. This is because the decision to not vote tends to be persistent between elections (see, for example, the true values in Table 7), and so we might assume that most of these voters would not have voted if not for the compulsory voting policy in place. In an extreme scenario, assuming that none of the runoff N-B-A voters would have voted, the Approve option would have won the election with approximately 54%. In a more realistic scenario, the perception of this election as more relevant than usual would have likely led to an increase in new voters relative to other elections, even with a voluntary voting policy. However, given that the bulk of the

---

<sup>16</sup>Despite the possibility of a fine, the probability of actually being fined is very low. The perceived likelihood of being fined is likely much higher than the actual probability, however, which possibly led to the significant increase in the number of votes.

<sup>17</sup>We pursue this issue further in Altman et al. (2023).

new votes for the option to reject came from N-B-A voters, it is highly likely that the result under a voluntary voting policy would have been the approval of the draft rather than its rejection.

## 8. Conclusion

This paper presented a novel approach to ecological inference, particularly to estimating voter transition matrices between two elections. Our estimation procedure has the benefit of having a clear intuition, which contrasts with the increasingly elaborate methods developed to deal with this problem. In the 2x2 case, we divide voters in each election into two groups and assume that voters in each group in the first election vote for each group in the second election with an unknown probability which varies across units. We assume that these probabilities are realizations of i.i.d. random variables with means and variances that vary across groups of voters in the first election. Furthermore, we assume that voters vote independently from each other. These assumptions imply that the conditional expected value of the number of votes for each group in the second election is a function of these probabilities. To estimate the means and variances of the previous random variables, we impose these first moment conditions as constraints. We do this by replacing the expected values with their sample means, in the spirit of the method of moments. Our estimates for the expected values of the unit-level probabilities then consist of the points in the unit square that minimize the squared Euclidean distance to these constraints.

We show that, under our assumptions, the resulting estimators are consistent. We also obtain consistent estimators for the variance of the unit-level values across units. Furthermore, our procedure has properties that simplify estimation significantly and make it less computationally intensive than current approaches. In particular, we obtain our estimates by solving an optimization problem with a unique solution and a quadratic objective function. This allows our model to conduct inference on large voter transition matrices and with large datasets, which is not always the case for the simulation-based methods in [King \(1997\)](#), [King, Rosen, and Tanner \(1999\)](#), and [Rosen et al. \(2001\)](#). Furthermore, simulation studies show that it accurately estimates the true parameter values in small samples. In these simulations, we find that the performance of our model (as measured by the RMSE of its estimates) is similar to more established approaches and that it provides superior estimates when the true parameter values are close to 0 and 1. Finally, we show how our model may be extended in a straightforward way to deal with the presence of multiple clusters in the true values for the voter transitions.

We illustrated our model with two 2x2 applications and one RxC application of ecological inference. In both 2x2 applications, the true values of the transition matrix were known, which allowed us to compare the estimates of our model to the true values and to the [King, Rosen, and Tanner \(1999\)](#) model. In our application using US voter registration data by race, our model returned similar estimates to more established approaches. We presented three estimates in our application to the 2013 Chilean presidential elections. We were able to show that our model is better suited than other approaches to work with large datasets, that it provides more accurate estimates of the true values, and that it can be extended in a simple way to work when there are multiple clusters of true voter transitions across units. Our approach for this scenario correctly identified the two clusters observed in the true values of the voter transitions and provided better, more precise estimates than those that resulted from estimating our model ignoring the presence of multiple clusters.

For our RxC application, we estimated the voter transition matrix between the 2021 Chilean presidential election runoff and the 2022 Constitutional Plebiscite, in which voters chose whether to approve or reject the draft for a new constitution proposed by the Constitutional Convention. Our estimates were similar to the [Rosen et al. \(2001\)](#) model, a more established approach to ecological inference. Furthermore, our results have interesting implications for the analysis of these elections. In particular, we find that the voters brought on to the 2022 plebiscite by new compulsory voting policies heavily favored the option to reject the draft. Specifically, close to 90% of voters who did not vote in the runoff but did vote in the 2022 plebiscite voted to reject it. Under the assumption that most of these voters would have chosen not to vote under a voluntary voting policy, this suggests that in the absence of a compulsory voting policy the result would have likely been the approval of the constitutional draft rather than its rejection. In the extreme scenario where only voters who voted for one of the competing candidates in the runoff voted in the 2022 plebiscite, the draft for a new constitution would have been approved by a margin of 54% to 46%.

The present work may be expanded in several directions. One possibility is to refine and emphasize the extension to multiple clusters. This is important since there are numerous potential applications where we might expect there to be more than one cluster in the true voter transitions, as was the case for La Florida in the 2x2 application in Chile. Furthermore, computationally intensive methods of ecological inference might not be as straightforward to extend to deal with this scenario, meaning our approach may offer further advantages to other models in cases with multiple clusters. Another avenue to develop this work further is to rigorously extend our estimator to the RxC case by proving

that the properties listed in Section 4 continue to hold in this more general case.

## References

- Altman, David, Juan Díaz, Eduardo Engel, and Benjamín Peña. 2023. “¿Voto obligatorio o péndulo? Explicando las (in)consistencias de la ciudadanía chilena.” *El País*.
- Amemiya, Takeshi. 1985. *Advanced Econometrics*.: Harvard University Press.
- Andreadis, Ioannis, and Theodore Chadjipadelis. 2009. “A Method for the Estimation of Voter Transition Rates.” *Journal of Elections, Public Opinion and Parties* 19 (2): 203–218.
- Brown, Philip J., and Clive D. Payne. 1986. “Aggregate Data, Ecological Regression, and Voting Transitions.” *Journal of the American Statistical Association* 81 (394): 452–460.
- Chattopadhyay, Ambarish, and José R Zubizarreta. 2022. “On the implied weights of linear regression for causal inference.” *Biometrika*. asac058.
- Duncan, Otis Dudley, and Beverly Davis. 1953. “An Alternative to Ecological Correlation.” *American Sociological Review* 18 (6): 665.
- Freedman, D. A., S. P. Klein, M. Ostland, and M. R. Roberts. 1998. *Journal of the American Statistical Association* 93 (444): 1518–1522.
- Freedman, D. A., M. Ostland, M. R. Roberts, and S. P. Klein. 1999. “Response to King’s Comment.” *Journal of the American Statistical Association* 94 (445): 355–357.
- Füle, Erika. 1994. “Estimating voter transitions by ecological regression.” *Electoral Studies* 13 (4): 313–330.
- Goodman, Leo A. 1953. “Ecological Regressions and Behavior of Individuals.” *American Sociological Review* 18 (6): 663.
- Goodman, Leo A. 1959. “Some Alternatives to Ecological Correlation.” *American Journal of Sociology* 64 (6): 610–625.
- Greene, William. 2018. *Econometric Analysis*. 8th ed.: Pearson.
- Greiner, and Kevin M. Quinn. 2009. “R x C Ecological Inference: Bounds, Correlations, Flexibility and Transparency of Assumptions.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 172 (1): 67–81.
- Imai, Kosuke, Ying Lu, and Aaron Strauss. 2008. “Bayesian and Likelihood Inference for  $2 \times 2$  Ecological Tables: An Incomplete-Data Approach.” *Political Analysis* 16 (1): 41–69.
- Kawai, Kei, and Yasutora Watanabe. 2013. “Inferring Strategic Voting.” *American Economic Review* 103 (2): 624–62.

- King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.*: Princeton University Press.
- King, Gary. 1999. “The Future of Ecological Inference Research: A Comment on Freedman et al.” *Journal of the American Statistical Association* 94 (445): 352–355.
- King, Gary, Ori Rosen, and Martin A Tanner. 1999. “Binomial-Beta Hierarchical Models for Ecological Inference.” *Sociological Methods & Research* 28 (1): 61–90.
- King, Gary, Ori Rosen, Martin Tanner, and Alexander F. Wagner. 2008. “Ordinary Economic Voting Behavior in the Extraordinary Election of Adolf Hitler.” *The Journal of Economic History* 68 (4): 951–996.
- King, Gary, Martin A Tanner, and Ori Rosen. 2004. *Ecological Inference: New Methodological Strategies.*: Cambridge University Press.
- Klima, André, Thomas Schlesinger, Paul W. Thurner, and Helmut Küchenhoff. 2019. “Combining Aggregate Data and Exit Polls for the Estimation of Voter Transitions.” *Sociological Methods & Research* 48 (2): 296–325.
- Klima, André, Paul W Thurner, Christoph Molnar, Thomas Schlesinger, and Helmut Küchenhoff. 2016. “Estimation of Voter Transitions Based on Ecological Inference: An Empirical Assessment of Different Approaches.” *AStA Advances in Statistical Analysis* 100 (2): 133–159.
- Knudson, Karin C., Gabe Schoenbach, and Amariah Becker. 2021. “PyEI: A Python package for ecological inference.” *Journal of Open Source Software* 6 (64): 3397.
- Lau, Olivia, Ryan T Moore, and Michael Kellermann. 2007. “eiPack: R  $\times$  C ecological inference and higher-dimension data management.” *New Functions for Multivariate Analysis* 7 (1): 43.
- Manski, Charles F. 2018. “Credible ecological inference for medical decisions with personalized risk assessment.” *Quantitative Economics* 9 (2): 541–569.
- McCarthy, Colm, and Terence M. Ryan. 1977. “Estimates of Voter Transition Probabilities from the British General Elections of 1974.” *Journal of the Royal Statistical Society: Series A (General)* 140 (1): 78–85.
- Núñez, Lucas. 2016. “Expressive and strategic behavior in legislative elections in Argentina.” *Political Behavior* 38 (4): 899–920.
- Pavía, Jose M., and Rafael Romero. 2022. “Improving Estimates Accuracy of Voter Transitions. Two New Algorithms for Ecological Inference Based on Linear Programming.” *Sociological Methods & Research* 0 (0): 00491241221092725.

- Plescia, Carolina, and Lorenzo De Sio. 2018. “An Evaluation of the Performance and Suitability of  $R \times C$  Methods for Ecological Inference with Known True Values.” *Quality & Quantity* 52 (2): 669–683.
- Pons, Vincent, and Clémence Tricaud. 2018. “Expressive Voting and Its Cost: Evidence From Runoffs With Two or Three Candidates.” *Econometrica* 86 (5): 1621–1649.
- Robinson, W. S. 1950. “Ecological Correlations and the Behavior of Individuals.” *American Sociological Review* 15 (3): 351–357.
- Rosen, Ori, Wenxin Jiang, Gary King, and Martin A Tanner. 2001. “Bayesian and Frequentist Inference for Ecological Inference: The  $R \times C$  Case.” *Statistica Neerlandica* 55 (2): 134–156.
- Upton, Graham J. G. 1978. “A Note on the Estimation of Voter Transition Probabilities.” *Journal of the Royal Statistical Society: Series A (General)* 141 (4): 507–512.
- Wakefield, Jon. 2004. “Ecological Inference for  $2 \times 2$  Tables.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 167 (3): 385–425.
- Zubizarreta, José R. 2015. “Stable weights that balance covariates for estimation with incomplete outcome data.” *Journal of the American Statistical Association* 110 (511): 910–922.

## Appendix A. Proof of Proposition 1

Take the first formulation of the problem:

$$\begin{aligned}
& \min_{\{\hat{p}, \hat{q}\} \in [0,1]^2} \sum_{i=1}^N [(\hat{p}_i(\hat{p}, \hat{q}) - \hat{p})^2 + (\hat{q}_i(\hat{p}, \hat{q}) - \hat{q})^2] \\
& \text{s.t. } (\hat{p}_i, \hat{q}_i) \in \underset{\hat{p}_i, \hat{q}_i \in [0,1]}{\text{argmin}} [(\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2] \quad \forall i = 1, \dots, N \\
& \text{s.t. } v_i = n_i \hat{p}_i + m_i \hat{q}_i
\end{aligned} \tag{A1}$$

Note that the objective function in the outer level of the problem is additive over  $i$  and simply add the problems that are solved in the inner level for each  $i$ . This means that solving the inner level of the problem of a pair  $\hat{p}_i$  and  $\hat{q}_i$  under its restrictions is the same as minimizing the objective function of the outer level choosing  $\hat{p}_i$  and  $\hat{q}_i$  under the same restrictions as the inner level. Therefore, we can turn this problem into a standard optimization problem in the following way:

$$\begin{aligned}
& \min_{\{\{\hat{p}, \hat{q}\}, \{\hat{p}_i, \hat{q}_i\}_{i=1}^N\}} \sum_{i=1}^N [(\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2] \\
& \text{s.t. } \hat{p}, \hat{q} \in [0, 1] \\
& \quad v_i = n_i \hat{p}_i + m_i \hat{q}_i, i = 1, 2, \dots, N \\
& \quad \hat{p}_i, \hat{q}_i \in [0, 1], i = 1, 2, \dots, N
\end{aligned} \tag{A2}$$

Intuitively, in this equation, we are simultaneously restricting the first moment constraints to the unit square and choosing the point  $(\hat{p}, \hat{q})$  that is closest to these constraints. Since we are restricting every  $\hat{p}_i$  and  $\hat{q}_i$  to be between 0 and 1, the bound constraints on  $\hat{p}$  and  $\hat{q}$  are redundant. Therefore, this simplifies to the following problem:

$$\begin{aligned}
& \min_{\{\{\hat{p}, \hat{q}\}, \{\hat{p}_i, \hat{q}_i\}_{i=1,2,\dots,N}\}} \sum_{i=1}^N [(\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2] \\
& \text{s.t. } v_i = n_i \hat{p}_i + m_i \hat{q}_i, i = 1, 2, \dots, N \\
& \quad \hat{p}_i, \hat{q}_i \in [0, 1], i = 1, 2, \dots, N
\end{aligned} \tag{A3}$$

And the solutions to this problem for  $\hat{p}$  and  $\hat{q}$  are known to be  $\hat{p} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i$  and  $\hat{q} = \frac{1}{N} \sum_{i=1}^N \hat{q}_i$ . Replacing this into the objective function, we get the second formulation of the problem.

## Appendix B. Proof of Proposition 2

To briefly recap the notation, consider  $N$  units indexed by  $i$ . For each unit, we have the following information:

- $n_i$ : Number of votes in the first election.
- $m_i$ : Number of people that did not vote in the first election.
- $v_i$ : Number of votes in the second election.

It is easier to solve this problem by introducing some extra notation. Rename  $\hat{p}_i$  to  $\hat{p}_{1i}$ , the probability that someone from group 1, first round voters, votes again, and  $\hat{q}_i$  to  $\hat{p}_{2i}$ , the probability that someone from group 2, first round non-voters, votes in the runoff. The problem that we solve to get the estimators for each unit is the following:

$$\begin{aligned} \min_{\{\hat{p}_i, \hat{q}_i\}_{i=1, \dots, N}} \quad & \sum_{i=1}^N \left[ \sum_{c=1}^2 (\hat{p}_{ci} - \frac{1}{N} \hat{p}_{ci})^2 \right] \\ \text{s.t.} \quad & n_i \hat{p}_{1i} + m_i \hat{p}_{2i} = v_i, \quad i = 1, \dots, N \\ & \hat{p}_{1i}, \hat{p}_{2i} \in [0, 1], \quad i = 1, \dots, N \end{aligned} \tag{B1}$$

To prove this is a convex optimization problem, we will first prove that the restrictions define a convex set in which we choose the parameter values. We then prove that the objective function is convex in this convex set. First, define the following sets for each restriction in each unit:

$$G_i = \{(\hat{p}_{1i}, \hat{p}_{2i}) \in [0, 1]^2 : n_i \hat{p}_{1i} + m_i \hat{p}_{2i} = v_i\}, \quad i = 1, \dots, N$$

It is straightforward to see that these sets are convex for all  $i$ . Taking points  $\mathbf{p}, \mathbf{p}' \in G_i$ ,  $\lambda \mathbf{p} + (1 - \lambda) \mathbf{p}' \in [0, 1]^2$  for  $\lambda \in [0, 1]$ , and

$$\begin{aligned} n_i[\lambda p_{1i} + (1 - \lambda) p'_{1i}] + m_i[\lambda p_{2i} + (1 - \lambda) p'_{2i}] &= \lambda(n_i p_{1i} + m_i p_{2i}) + (1 - \lambda)(n_i p'_{1i} + m_i p'_{2i}) \\ &= \lambda v_i + (1 - \lambda) v_i = v_i \end{aligned}$$

given that  $n_i p_{1i} + m_i p_{2i} = v_i$  y  $n_i p'_{1i} + m_i p'_{2i} = v_i$ . The set  $G$  over which we minimize the objective function is the intersection of all sets  $G_i$ . Given that each set is convex and the arbitrary intersection of convex sets is also convex, the restrictions define a convex set  $G$

over which we minimize the objective function.

Given that  $G$  is a convex set, to prove that the objective function  $F$  is convex, it is enough to prove that the following is true for  $\mathbf{p}, \mathbf{p}' \in G$ :

$$F(\lambda\mathbf{p} + (1 - \lambda)\mathbf{p}') \leq \lambda F(\mathbf{p}) + (1 - \lambda)F(\mathbf{p}')$$

We first rearrange the left-hand side:

$$\begin{aligned} F(\lambda\mathbf{p} + (1 - \lambda)\mathbf{p}') &= \sum_{i=1}^N \left[ \sum_{c=1}^2 \left( (\lambda p_{ci} + (1 - \lambda)p'_{ci}) - \frac{1}{N} \sum_{i=1}^N (\lambda p_{ci} + (1 - \lambda)p'_{ci}) \right)^2 \right] \\ &= \sum_{i=1}^N \left[ \sum_{c=1}^2 \left( \lambda(p_{ci} - \frac{1}{N} \sum_{i=1}^N p_{ci}) + (1 - \lambda)(p'_{ci} - \frac{1}{N} \sum_{i=1}^N p'_{ci}) \right)^2 \right] \end{aligned}$$

Then the right hand side:

$$\begin{aligned} \lambda F(\mathbf{p}) + (1 - \lambda)F(\mathbf{p}') &= \lambda \sum_{i=1}^N \left[ \sum_{c=1}^2 (p_{ci} - \frac{1}{N} \sum_{i=1}^N p_{ci})^2 \right] + (1 - \lambda) \sum_{i=1}^N \left[ \sum_{c=1}^2 (p'_{ci} - \frac{1}{N} \sum_{i=1}^N p'_{ci})^2 \right] \\ &= \sum_{i=1}^N \left[ \sum_{c=1}^2 \left( \lambda(p_{ci} - \frac{1}{N} \sum_{i=1}^N p_{ci})^2 + (1 - \lambda)(p'_{ci} - \frac{1}{N} \sum_{i=1}^N p'_{ci})^2 \right) \right] \end{aligned}$$

Finally, and given that  $f(x) = x^2$  is a strictly convex function, Jensen's inequality implies that the following is true for all  $r$  and  $c$ :

$$\begin{aligned} \left( \lambda(p_{ci} - \frac{1}{N} \sum_{i=1}^N p_{ci}) + (1 - \lambda)(p'_{ci} - \frac{1}{N} \sum_{i=1}^N p'_{ci}) \right)^2 &< \\ &\left( \lambda(p_{ci} - \frac{1}{N} \sum_{i=1}^N p_{ci})^2 + (1 - \lambda)(p'_{ci} - \frac{1}{N} \sum_{i=1}^N p'_{ci})^2 \right) \end{aligned}$$

Then, we have that  $F(\lambda\mathbf{p} + (1 - \lambda)\mathbf{p}') < \lambda F(\mathbf{p}) + (1 - \lambda)F(\mathbf{p}')$ , and  $F$  is a strictly convex function. Since strict convexity implies convexity, it follows that (??) is an optimization problem where we minimize a convex function over a convex set. That is, it is a convex optimization problem. Furthermore, given that  $F$  is strictly convex, it has a unique global optimum. That it is quadratic is clear by looking at the objective function in the second formulation of the problem.

## Appendix C. Proof of Proposition 3

### Notation

For each area  $i = 1, 2, \dots, n$ , define the following variables:

- $n_i$ : Number of people who voted in the first election.
- $m_i$ : Number of people who did not vote in the first election.
- $v_i$ : Number of votes in the second election.
- $p_i$ : Probability that someone who voted in the first round votes again in unit  $i$ .

If we condition on  $p_i$  and  $q_i$ , the number of votes in the second election is distributed as the sum of two independent Binomial random variables:

$$(v_i | p_i, q_i) \sim \text{Bin}(n_i, p_i) + \text{Bin}(m_i, q_i),$$

where

$$\begin{aligned} p_i &\sim (\mu_p, \sigma_p^2), \\ q_i &\sim (\mu_q, \sigma_q^2). \end{aligned}$$

and both are bounded between 0 and 1<sup>18</sup>. Then, using the law of iterated expectations and the law of total variance:

$$\begin{aligned} \mathbb{E}(v_i) &= n_i \mu_p + m_i \mu_q \\ \text{Var}(v_i) &= n_i [\mu_p(1 - \mu_p) + \sigma_p^2(n_i - 1)] + m_i [\mu_q(1 - \mu_q) + \sigma_q^2(m_i - 1)] \end{aligned}$$

We can then write  $v_i$  as follows:

$$v_i = n_i \mu_p + m_i \mu_q + \varepsilon_i$$

where  $\varepsilon_i = (v_i - \mathbb{E}(v_i))$ , and the following holds:

$$\begin{aligned} \mu_i &\equiv \mathbb{E}(\varepsilon_i) = 0 \\ \sigma_i^2 &\equiv \text{Var}(\varepsilon_i) = n_i [\mu_p(1 - \mu_p) + \sigma_p^2(n_i - 1)] + m_i [\mu_q(1 - \mu_q) + \sigma_q^2(m_i - 1)] \end{aligned}$$

---

<sup>18</sup>Note that these are the parameters, not our estimates.

For ease of notation, define  $w_i = (n_i^2 + m_i^2)^{-1}$ , as well as the following matrices:

$$\begin{aligned}\boldsymbol{\beta} &= (\mu_p \quad \mu_q)' \\ \mathbf{Y} &= (v_1 \quad v_2 \quad \dots \quad v_N)' \\ \mathbf{X} &= \begin{pmatrix} n_1 & n_2 & \dots & n_N \\ m_1 & m_2 & \dots & m_N \end{pmatrix}' \\ \boldsymbol{\varepsilon} &= (\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_N)' \\ \mathbf{W} &= \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_N \end{pmatrix}\end{aligned}$$

we can then write our model as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

We can rewrite our vector of estimators  $\hat{\boldsymbol{\beta}}$  when there are only interior solutions using the previous matrices as follows:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \sum_{i=1}^N n_i^2 w_i & \sum_{i=1}^N n_i m_i w_i \\ \sum_{i=1}^N n_i m_i w_i & \sum_{i=1}^N m_i^2 w_i \end{bmatrix}^{-1} \frac{1}{N} \begin{bmatrix} \sum_{i=1}^N n_i w_i v_i \\ \sum_{i=1}^N m_i w_i v_i \end{bmatrix} \quad (\text{C1})$$

$$= (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}\mathbf{Y}) \quad (\text{C2})$$

$$(\text{C3})$$

while the estimators for the fractions within each unit are

$$\hat{p}_i = \gamma_{1i}\hat{p} - \gamma_{2i}\hat{q} + n_i w_i v_i \quad (\text{C4})$$

$$\hat{q}_i = \gamma_{3i}\hat{q} - \gamma_{2i}\hat{p} + m_i w_i v_i \quad (\text{C5})$$

where

$$\gamma_{1i} = n_i^2 w_i$$

$$\gamma_{2i} = n_i m_i w_i$$

$$\gamma_{3i} = m_i^2 w_i$$

and  $\gamma_j = \frac{1}{N} \sum_{i=1}^N \gamma_{ji}$  for  $j = 1, 2, 3$ . Then, if we ignore bound constraints, we have a particular case of an OLS estimation with heteroskedasticity. We can rewrite  $\hat{\boldsymbol{\beta}}$  as follows:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \mathbf{X} \right)^{-1} \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \boldsymbol{\varepsilon} \right)$$

We can also get closed expressions for  $\hat{p}$ ,  $\hat{q}$ ,  $\hat{p}_i$ , and  $\hat{q}_i$  in the case with interior solutions by solving (C3) and plugging these solutions into (C4) and (C5). This results in the following expressions:

$$\begin{aligned} \hat{p} &= \frac{1}{N} \sum_{i=1}^N (\Gamma_3 n_i - \Gamma_2 m_i) w_i v_i \\ \hat{q} &= \frac{1}{N} \sum_{i=1}^N (\Gamma_1 m_i - \Gamma_2 n_i) w_i v_i \\ \hat{p}_i &= \frac{1}{N} \sum_j [(\gamma_{3j} \Gamma_3 + \gamma_{2j} \Gamma_2) n_j - (\gamma_{3j} \Gamma_2 + \gamma_{2j} \Gamma_1) m_j] w_j v_j + n_i w_i v_i \\ \hat{q}_i &= \frac{1}{N} \sum_j [(\gamma_{1j} \Gamma_1 + \gamma_{2j} \Gamma_2) m_j - (\gamma_{1j} \Gamma_2 + \gamma_{2j} \Gamma_3) n_j] w_j v_j + m_i w_i v_i \end{aligned}$$

where  $\Gamma_j = \frac{\gamma_j}{\gamma_1 \gamma_3 - \gamma_2^2}$ , for  $j = 1, 2, 3$ .

## Consistency

We know that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}' \mathbf{W} \boldsymbol{\varepsilon}).$$

From this, it follows that

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N} \left( \boldsymbol{\beta}, \frac{1}{N} \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \mathbf{X} \right)^{-1} \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \boldsymbol{\Omega} \mathbf{W}' \mathbf{X} \right) \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \mathbf{X} \right)^{-1} \right).$$

where

$$\boldsymbol{\Omega} = \mathbb{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}') = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix}.$$

Since it is unbiased,  $\hat{\beta}$  will converge in mean square to  $\beta$  if the variance goes to zero. For this to happen, we need  $\frac{1}{N}\mathbf{X}'\mathbf{W}\mathbf{X}$  and  $\frac{1}{N}\mathbf{X}'\mathbf{W}\mathbf{\Omega}\mathbf{W}'\mathbf{X}$  to converge to positive definite matrices as  $N \rightarrow \infty$ . If we let  $\tilde{\mathbf{X}} = \mathbf{W}^{\frac{1}{2}}\mathbf{X}$  and  $\tilde{\mathbf{\Omega}} = \mathbf{W}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{W}^{\frac{1}{2}}$ , we can write

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{1}{N} \left(\frac{1}{N}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} \left(\frac{1}{N}\tilde{\mathbf{X}}'\tilde{\mathbf{\Omega}}\tilde{\mathbf{X}}\right) \left(\frac{1}{N}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\right).$$

So we need to prove that the terms  $\frac{1}{N}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{X}}'\tilde{\mathbf{\Omega}}\tilde{\mathbf{X}}$  converge to positive definite matrices. To do this, it is sufficient to prove that the matrix  $\tilde{\mathbf{X}}$  meets the Grenander conditions and that the terms on the diagonal of  $\tilde{\mathbf{\Omega}}$ ,  $\sigma_i^2 w_i$ , are finite for all  $i$  (see [Amemiya 1985](#); [Greene 2018](#)). Letting  $\lambda_i = \frac{n_i^2}{m_i^2}$ ,  $a_i = \frac{1}{1+\lambda_i}$ , and  $b_i = \frac{1}{1+\lambda_i}$ , the Grenander conditions for  $\tilde{\mathbf{X}}$  are as follows:

- $\lim_{N \rightarrow \infty} \sum a_i = +\infty$  and  $\lim_{N \rightarrow \infty} \sum b_i = +\infty$ . For this to be true, we need  $1/\lambda_i$  to be bounded between 0 and any positive constant for all units or that as  $N \rightarrow \infty$ , there are always units in which  $1/\lambda_i$  is bounded between 0 and any positive constant. This is not restrictive in our context since it simply means that neither  $n_i$  nor  $m_i$  tend to zero for all units as  $N$  increases.
- $\lim_{N \rightarrow \infty} \frac{n_i^2}{\sum a_j} = 0$  and  $\lim_{N \rightarrow \infty} \frac{m_i^2}{\sum b_j} = 0$  for all  $i$ . Given that  $n_i$  and  $m_i$  are finite for all  $i$ , this holds as long as the previous condition holds.
- The full rank condition for the matrix  $\tilde{\mathbf{X}}$  is always met as the sample size increases. The only requirement for this to be true is that  $n_i \neq m_i$  for at least one unit  $i$  and that there is no constant  $c$  such that  $n_i = cm_i$  for all units.

For the condition involving  $\tilde{\mathbf{\Omega}}$ , note that

$$\begin{aligned} \sigma_i^2 w_i &= \frac{n_i[\mu_p(1 - \mu_p) + \sigma_p^2(n_i - 1)] + m_i[\mu_q(1 - \mu_q) + \sigma_q^2(m_i - 1)]}{n_i^2 + m_i^2} \\ &< \frac{n_i[1 + \sigma_p^2(n_i - 1)] + m_i[1 + \sigma_q^2(m_i - 1)]}{n_i^2 + m_i^2} \\ &< \frac{n_i[1 + n_i - 1] + m_i[1 + m_i - 1]}{n_i^2 + m_i^2} = 1 \end{aligned}$$

where we used that  $\mu_p(1 - \mu_p) < 1$  and  $\mu_q(1 - \mu_q) < 1$  in the first inequality and the fact that  $\sigma_p^2 < 1$  and  $\sigma_q^2 < 1$  in the second inequality. The latter inequalities result from Popoviciu's inequality on variances and the fact that each  $p_i$  and  $q_i$  is bounded between 0

and 1. Given the previous points, we know that both  $\frac{1}{N}\mathbf{X}'\mathbf{W}\mathbf{X}$  and  $\frac{1}{N}\tilde{\mathbf{X}}'\tilde{\Omega}\tilde{\mathbf{X}}$  converge to positive definite matrices as  $N$  increases. Therefore, the variance of  $\hat{\beta}$  goes to zero as the sample size increases, and  $\hat{\beta}$  converges in mean square to  $\beta$ . This implies that it converges in probability, which is the result stated in the main text.

### Asymptotic Normality

We first rewrite  $\hat{\beta}$ :

$$\sqrt{N}(\hat{\beta} - \beta) = \left( \frac{1}{N}\mathbf{X}'\mathbf{W}\mathbf{X} \right)^{-1} \left( \frac{1}{\sqrt{N}}\mathbf{X}'\mathbf{W}\boldsymbol{\varepsilon} \right)$$

We know that  $\frac{1}{N}\mathbf{X}'\mathbf{W}\mathbf{X}$  converges to a positive definite matrix, which we call  $\mathbf{Q}$ . Then, it is enough to know the distribution of the following vector:

$$\mathbf{v}_N = \mathbf{Q}^{-1} \frac{1}{\sqrt{N}}\mathbf{X}'\mathbf{W}\boldsymbol{\varepsilon} = \mathbf{Q}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{x}_i w_i \varepsilon_i = \mathbf{Q}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} n_i w_i \varepsilon_i \\ m_i w_i \varepsilon_i \end{pmatrix}$$

We use the Lindeberg-Feller CLT, for which we need to verify that some additional conditions are met. In our context, it is enough to prove that the variance of  $\sqrt{w_i}\varepsilon_i$  is finite for all  $i$ , that the average variance converges to a finite constant, and that the average variance is not dominated by any term (see [Greene 2018](#)). Given that the variance of the previous term is  $\sigma_i^2 w_i$ , we can summarize the previous points in the following requirements:

- $\sigma_i^2 w_i < \infty$
- $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2 w_i < \infty$
- $\lim_{N \rightarrow \infty} \frac{\max(\sigma_i^2 w_i)}{\sum_{i=1}^N \sigma_i^2 w_i} = 0$

We have already shown that the first condition holds. The second condition holds given that the sum is a series of non-negative terms with partial sums bounded by 1, which holds since  $\sigma_i^2 w_i < 1$ . Therefore, it converges to a positive constant. The third point holds given that the denominator tends to infinity as  $N$  grows and each  $\sigma_i^2 w_i$  is finite. We can then use the Lindeberg-Feller CLT to get the following result:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N} \left( \mathbf{0}, \mathbf{Q}^{-1} \text{plim} \left( \frac{1}{N}\mathbf{X}'\mathbf{W}\Omega\mathbf{W}'\mathbf{X} \right) \mathbf{Q}^{-1} \right)$$

which is the same result stated in the main text.

## Appendix D. Proof of Corollary 1

Start with the following equation:

$$\begin{aligned}
& \min_{\{\hat{p}_i, \hat{q}_i\}_{i=1, \dots, N}} \sum_{i=1}^N \left[ (\hat{p}_i - \frac{1}{N} \sum_{i=1}^N \hat{p}_i)^2 + (\hat{q}_i - \frac{1}{N} \sum_{i=1}^N \hat{q}_i)^2 \right] \\
& \text{s.t.} \quad n_i \hat{p}_i + m_i \hat{q}_i = v_i, \quad i = 1, \dots, N \\
& \quad \hat{p}_i, \hat{q}_i \in [0, 1], \quad i = 1, \dots, N
\end{aligned} \tag{D1}$$

Our estimators for  $\mu_p$  and  $\mu_q$  will be  $\frac{1}{N} \sum_{i=1}^N \hat{p}_i$  and  $\frac{1}{N} \sum_{i=1}^N \hat{q}_i$ , respectively. We want to prove that  $\hat{p}$  and  $\hat{q}$  are consistent when imposing bound constraints on  $\hat{p}_i$  and  $\hat{q}_i$  in (D1). We present the proof for  $\hat{p}$ . The proof for  $\hat{q}$  is analogous.

Let  $\hat{p}_i^{UR}$  be the estimators that result from (D1) without using bound constraints and let  $\hat{p}^{UR} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i^{UR}$ . Define  $\hat{p}_i^{BC}$  and  $\hat{p}^{BC}$  similarly for the case with bound constraints. Note that

$$\begin{aligned}
\text{MSE}(\hat{p}^{BC}) &= \mathbb{E} \left[ (\hat{p}^{BC} - \mu_p)^2 \right] \\
&= \mathbb{E} \left[ \left( \frac{1}{N} \sum_{i=1}^N \hat{p}_i^{BC} - \mu_p \right)^2 \right] \\
&= \mathbb{E} \left[ \left( \frac{1}{N} \sum_{i=1}^N (\hat{p}_i^{BC} - \mu_p) \right)^2 \right] \\
&\leq \mathbb{E} \left[ \left( \frac{1}{N} \sum_{i=1}^N (\hat{p}_i^{UR} - \mu_p) \right)^2 \right] = \text{MSE}(\hat{p}^{UR})
\end{aligned}$$

The inequality comes by noting that since  $0 \leq \mu_p \leq 1$ , the difference between each  $\hat{p}_i^{BC}$  and  $\mu_p$  will be equal to the difference between  $\hat{p}_i^{UR}$  and  $\mu_p$  if and only if  $\hat{p}_i^{UR}$  falls between 0 and 1. Otherwise,  $\hat{p}_i^{BC}$  will be closer to  $\mu_p$  than  $\hat{p}_i^{UR}$ . We can then write

$$0 \leq \text{MSE}(\hat{p}^{BC}) \leq \text{MSE}(\hat{p}^{UR})$$

Since  $\hat{p}^{UR}$  is mean square consistent by Proposition 1 and Proposition 3, letting  $N \rightarrow \infty$  the term on the right-hand side goes to zero. The squeeze theorem then implies that  $\hat{p}^{BC}$  is mean square consistent as well.

## Appendix E. Proof of Corollary 2

Start with the following expression:

$$\frac{1}{N} \sum \left( \hat{p}_i - \frac{1}{N} \sum \hat{p}_i \right)^2 + \frac{1}{N} \sum \left( \hat{q}_i - \frac{1}{N} \sum \hat{q}_i \right)^2.$$

This is the same as

$$\frac{1}{N} \sum (\hat{p}_i - \hat{p})^2 + \frac{1}{N} \sum (\hat{q}_i - \hat{q})^2.$$

We can replace  $\hat{p}_i$  and  $\hat{q}_i$  by their known values to arrive at

$$\frac{1}{N} \sum n_i^2 w_i^2 \hat{\varepsilon}_i^2 + \frac{1}{N} \sum m_i^2 w_i^2 \hat{\varepsilon}_i^2,$$

where  $\hat{\varepsilon}_i = v_i - n_i \hat{p} - m_i \hat{q}$ . We then reorder and note the following:

$$\begin{pmatrix} \frac{1}{N} \sum n_i^2 w_i^2 \hat{\varepsilon}_i^2 \\ \frac{1}{N} \sum m_i^2 w_i^2 \hat{\varepsilon}_i^2 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \frac{1}{N} \sum n_i^2 w_i^2 \sigma_i^2 \\ \frac{1}{N} \sum m_i^2 w_i^2 \sigma_i^2 \end{pmatrix} = \text{diag} \left( \frac{1}{N} \mathbf{X}' \mathbf{W} \mathbf{\Omega} \mathbf{W}' \mathbf{X} \right)$$

where we slightly abuse notation and let  $\text{diag}(\mathbf{Z})$  denote a column vector with the elements on the diagonal of matrix  $\mathbf{Z}$ , and  $(\frac{1}{N} \mathbf{X}' \mathbf{W} \mathbf{\Omega} \mathbf{W}' \mathbf{X})$  is defined in Appendix C, where we also proved that it converges to a positive definite matrix when proving the asymptotic normality of the estimators without bound constraints. The matrix on the right-hand side is therefore well-defined and equal to the following expression:

$$\begin{aligned} \begin{pmatrix} \frac{1}{N} \sum n_i^2 w_i^2 \sigma_i^2 \\ \frac{1}{N} \sum m_i^2 w_i^2 \sigma_i^2 \end{pmatrix} &= \frac{1}{N} \begin{pmatrix} \sum n_i^3 w_i^2 & \sum n_i^2 m_i w_i^2 \\ \sum n_i m_i^2 w_i^2 & \sum m_i^3 w_i^2 \end{pmatrix} \begin{pmatrix} p(1-p) \\ q(1-q) \end{pmatrix} \\ &+ \frac{1}{N} \begin{pmatrix} \sum n_i^3 (n_i - 1) w_i^2 & \sum n_i^2 m_i (m_i - 1) w_i^2 \\ \sum n_i (n_i - 1) m_i^2 w_i^2 & \sum m_i^3 (m_i - 1) w_i^2 \end{pmatrix} \begin{pmatrix} \sigma_p^2 \\ \sigma_q^2 \end{pmatrix} \end{aligned}$$

Then:

$$\begin{aligned} \frac{1}{N} \begin{pmatrix} \sum (\hat{p}_i - \hat{p})^2 \\ \sum (\hat{q}_i - \hat{q})^2 \end{pmatrix} &\xrightarrow{p} \frac{1}{N} \begin{pmatrix} \sum n_i^3 w_i^2 & \sum n_i^2 m_i w_i^2 \\ \sum n_i m_i^2 w_i^2 & \sum m_i^3 w_i^2 \end{pmatrix} \begin{pmatrix} p(1-p) \\ q(1-q) \end{pmatrix} \\ &+ \frac{1}{N} \begin{pmatrix} \sum n_i^3 (n_i - 1) w_i^2 & \sum n_i^2 m_i (m_i - 1) w_i^2 \\ \sum n_i (n_i - 1) m_i^2 w_i^2 & \sum m_i^3 (m_i - 1) w_i^2 \end{pmatrix} \begin{pmatrix} \sigma_p^2 \\ \sigma_q^2 \end{pmatrix}. \end{aligned} \tag{E1}$$

Solving for  $\sigma_p^2$  and  $\sigma_q^2$  and replacing  $\mu_p$  and  $\mu_q$  by their estimators, we can then estimate  $\sigma_p^2$  and  $\sigma_q^2$  as follows:

$$\begin{bmatrix} \hat{\sigma}_p^2 \\ \hat{\sigma}_q^2 \end{bmatrix} = N \begin{bmatrix} \sum n_i^3(n_i - 1)w_i^2 & \sum n_i^2m_i(m_i - 1)w_i^2 \\ \sum n_i(n_i - 1)m_i^2w_i^2 & \sum m_i^3(m_i - 1)w_i^2 \end{bmatrix}^{-1} \\ \left[ \frac{1}{N} \begin{bmatrix} \sum (\hat{p}_i - \hat{p})^2 \\ \sum (\hat{q}_i - \hat{q})^2 \end{bmatrix} - \frac{1}{N} \begin{bmatrix} \sum n_i^3w_i^2 & \sum n_i^2m_iw_i^2 \\ \sum n_im_i^2w_i^2 & \sum m_i^3w_i^2 \end{bmatrix} \begin{bmatrix} \hat{p}(1 - \hat{p}) \\ \hat{q}(1 - \hat{q}) \end{bmatrix} \right]$$

These estimators are consistent by (E1), the result of Proposition 3 on the consistency of  $\hat{p}$  and  $\hat{q}$ , and the continuous mapping theorem.

## Appendix F. Proof of Proposition 4

Start by taking the first formulation of the problem:

$$\begin{aligned}
& \min_{\{\hat{p}, \hat{q}\}} \sum_{i=1}^N \psi_i [(\hat{p}_i(\hat{p}, \hat{q}) - \hat{p})^2 + (\hat{q}_i(\hat{p}, \hat{q}) - \hat{q})^2] \\
& \text{s.t. } (\hat{p}_i, \hat{q}_i) \in \underset{\hat{p}_i, \hat{q}_i \in [0,1]}{\operatorname{argmin}} [(\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2] \quad \forall i = 1, \dots, N \quad (\text{F1}) \\
& \text{s.t. } v_i = n_i \hat{p}_i + m_i \hat{q}_i
\end{aligned}$$

As in the proof for Proposition 1, since the objective function in the outer level of the problem is additive over  $i$ , solving the inner level of the problem of a pair  $\hat{p}_i$  and  $\hat{q}_i$  under its restrictions is the same as minimizing the objective function of the outer level choosing  $\hat{p}_i$  and  $\hat{q}_i$  under the same restrictions as the inner level. Therefore, we can turn this problem into a standard optimization problem in the following way:

$$\begin{aligned}
& \min_{\{\{\hat{p}, \hat{q}\}, \{\hat{p}_i, \hat{q}_i\}_{i=1}^N\}} \sum_{i=1}^N [\psi_i (\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2] \\
& \text{s.t. } \hat{p}, \hat{q} \in [0, 1] \quad (\text{F2}) \\
& v_i = n_i \hat{p}_i + m_i \hat{q}_i, i = 1, 2, \dots, N \\
& \hat{p}_i, \hat{q}_i \in [0, 1], i = 1, 2, \dots, N
\end{aligned}$$

Again, the bound constraints on  $\hat{p}$  and  $\hat{q}$  are redundant given the bound constraints on each  $\hat{p}_i$  and  $\hat{q}_i$ . Therefore, this simplifies to the following problem:

$$\begin{aligned}
& \min_{\{\{\hat{p}, \hat{q}\}, \{\hat{p}_i, \hat{q}_i\}_{i=1,2,\dots,N}\}} \sum_{i=1}^N [\psi_i (\hat{p}_i - \hat{p})^2 + (\hat{q}_i - \hat{q})^2] \\
& \text{s.t. } v_i = n_i \hat{p}_i + m_i \hat{q}_i, i = 1, 2, \dots, N \quad (\text{F3}) \\
& \hat{p}_i, \hat{q}_i \in [0, 1], i = 1, 2, \dots, N
\end{aligned}$$

And the solutions to this problem for  $\hat{p}$  and  $\hat{q}$  are clearly  $\hat{p} = \sum_{i=1}^N \psi_i \hat{p}_i$  and  $\hat{q} = \sum_{i=1}^N \psi_i \hat{q}_i$ . Replacing this into the objective function, we get the second formulation of the problem. The proof for Proposition 2 for the case with weights follows by starting with the second formulation of the problem, letting  $p_{rc} = \sum_{i=1}^N \psi_i p_{rci}$ , and proceeding in the same manner outlined in Appendix B. The proof for Proposition 3 follows in the same manner as the proof Appendix C with only minimal changes by letting each  $w_i$  be  $\psi_i w_i$ , where  $w_i = (n_i^2 + m_i^2)^{-1}$ . The proofs for Corollary 1 and Corollary 2 are likewise unchanged if we

also let  $\hat{p} = \sum_{i=1}^N \psi_i \hat{p}_i$  and  $\hat{q} = \sum_{i=1}^N \psi_i \hat{q}_i$ .