UC Berkeley
CUDARE Working Papers

Title
An Information Theoretic Approach to Ecological Estimation and Inference

Permalink
https://escholarship.org/uc/item/7h03r00q

Authors
Judge, George G.
Miller, Douglas J.
Cho, Wendy K. T.

Publication Date
2003

An Information Theoretic Approach to Ecological Estimation and Inference[*]

George G. Judge[†]
University of California, Berkeley

Douglas J. Miller
Purdue University

Wendy K. Tam Cho
University of Illinois

[†] *Corresponding author*: 207 Giannini Hall, University of California, Berkeley, CA 94720–3310

CHAPTER 1

An Information Theoretic Approach to Ecological Estimation and Inference

George Judge
University of California, Berkeley

Douglas J. Miller
Purdue University

Wendy K. Tam Cho
University of Illinois at Urbana-Champaign

1.1 Introduction

In the social sciences, much of the data used for estimation and inference are available only in the form of averages or aggregate outcomes. Given this type of data restriction, researchers often use probabilities to represent information concerning the unknown and unobservable parameters of the underlying decision process. As a case in point, political scientists often face the question of how to process and recover information concerning voter behavior from precinct- or district-level data. These data are in many cases limited to aggregate vote counts, and individual specific information about voters in a precinct is seldom available.

Efforts to recover micro information from aggregate data generally result in ill-posed inverse problems which yield a multitude of feasible "solutions" due to the lack of sufficient information. In other words, ill-posed problems are fundamentally indeterminate because there are more unknowns than data points. Consequently, there is not enough information available from the data to uniquely solve the problem using traditional rules of logic. Seminal developments for coping with this long standing methodological challenge include Robinson (1950), Goodman (1953, 1959), Duncan and Davis (1953), Freedman, Klein, Sacks, Smyth, and Everett (1991), Achen and Shively (1995), and King (1997). Ill-posed inverse problems are not unique to political science and the literature is littered with pos-

sible solutions to related estimation and inference problems in economics and other fields (see for example Golan, Judge, and Miller, 1996).

Given the existence and importance of ecological estimation and inference problems in political science, we propose information theoretic procedures to recover estimates of the unknown conditional probabilities used as a basis for understanding voter behavior. In these problems, it is often possible to select feasible solutions or estimates that conform to the observed data, but the question lurking in the background is "what do these estimates mean or what question are they answering?"

Because the information theoretic and other formulations are based entirely on aggregate data that are limited, partial, and incomplete, the recovered conditional probabilities may not be appropriate for answering a range of important voter behavior questions. Hence, to make efficient use of aggregate election data, we must find some way to introduce additional structure into the modeling and information recovery process. One way to proceed is to specify a conceptual framework that provides a plausible basis for the underlying data generation process. Toward this end, in the second part of the paper, we suggest moment-based formulations that exploit the theoretical underpinnings of voter behavior and introduce important behavior parameters that facilitate the presentation and interpretation of the results. One purpose for adding this information or model structure is to provide a basis for converting a fundamentally ill-posed inverse problem into a well-posed problem. By reformulating the problem, we recover information at the appropriate level of aggregation on important voter response parameters along with the unknown conditional probabilities. The resulting formulations are semiparametric in the sense that the joint distribution of the underlying data is unspecified, apart from a finite set of moment conditions. These components form the basis for recovering the unknown response parameters and corresponding conditional probabilities and are used as standard operational tools in econometric information processing and recovery problems (Mittelhammer, Judge, and Miller, 2000). There certainly are many possible ways to approach ill-posed problems, so we emphasize that one must proceed cautiously when considering the significance of the estimates.

The paper proceeds as follows: In Section 1.2, we develop the notation consistent with the basic problem and develop a corresponding basis for modeling the aggregate data that focuses on the unknown conditional probabilities. In Section 1.3, we model the data as both a pure and a noisy inverse problem, suggest a solution, and interpret the recovered conditional probabilities. In Section 1.4, we suggest moment-based formulations that exploit the theoretical underpinnings of voter behavior and introduce important behavior parameters that facilitate the presentation and interpretation of the results. In Section 1.5 , we discuss the implications of the models and the proposed solutions as a basis for learning about voter behavior. Some examples based on real and synthetic data are presented in the Appendix.

1.2 Notation and Basic Inverse Model

To develop a model that will reflect the characteristics of voter response, consider the observed outcomes for a particular election across $i = 1, \ldots, m$ electoral units (e.g., precincts or districts). Each unit has $j = 1, \ldots, g$ types of individual voters and $k = 1, \ldots, c$ vote choices (e.g. candidates for office or propositions, including perhaps an abstention or no-vote category). For convenience and without loss of generality, we will adopt a framework where the election units are precincts and the vote choice is a set of candidates. For each precinct, the observed data are the number of votes for each candidate, $N_{i \cdot k} = \sum_{j=1}^{g} N_{ijk}$, and the number of voters in each group, $N_{ij \cdot} = \sum_{k=1}^{c} N_{ijk}$. The total number of ballots cast in the precinct is $N_i = \sum_{j=1}^{g} \sum_{k=1}^{c} N_{ijk}$. For any secret ballot, the total number of votes cast by each group for particular candidates in the election is unknown and *unobserved*. Given the observed data, our initial objective is to formulate an inverse model that will permit us to estimate N_{ijk}, the unobserved number of votes cast in precinct i by voters of type j for candidate k, from the sample of voters who voted in the election.

For the purposes of formulating the basic inverse model, the data may be stated in terms of the observed row or column proportions, i.e., for precinct i, $n_{i \cdot k} = N_{i \cdot k}/N_i$ or $n_{ij \cdot} = N_{ij \cdot}/N_i$. The inverse problem may be equivalently stated in terms of the proportion of voters in each category, $\beta_{ijk} = N_{ijk}/N_{ij \cdot} = n_{ijk}/n_{ij \cdot}$, where $\sum_{k=1}^{c} \beta_{ijk} = 1$ for each i and j. In this context, β_{ijk} may be interpreted as the conditional probability that voters in precinct i and group j voted for candidate k, where the conditioning indices are i and j. For example, in a study of split-ticket voting, the index j may represent votes for each of g national candidates from different parties, and the index k may represent the local candidates. The objective in this case would be to estimate the conditional probability that voters selected candidate k in the local election given that they voted for candidate j in the national election. In another application such as a study of polarized voting, the conditioning index j may represent characteristics of the electorate such as race or gender.

1.2.1 Modeling Voting Behavior as an Inverse Problem

The components of this information recovery problem for a particular precinct (i suppressed) are summarized in Table 1.1. The observed number of ballots cast by registered voters in each group ($N_{j \cdot}$) are the row sums, and the observed number of votes received by each candidate ($N_{\cdot k}$) are the column sums. What we do not know and cannot observe is the number of votes cast by each group, N_{jk}, or the proportion of votes cast by each group for each candidate, n_{jk}. If the conditional probabilities β_{jk} were known, we could derive the unknown number of voters as

$N_{jk} = \beta_{jk} N_{j\cdot}$. However, the conditional probabilities are unobserved and not accessible by direct measurement. Thus, we are faced with the cross-level inference problem, an inverse problem where we must use indirect, partial, and incomplete macro measurements as a basis for recovering the unknown conditional probabilities. The probability space interpretation gives the problem some minimal structure and provides a basis for learning from the data in a highly ambiguous situation.

The symbols in Table 1.1 and the corresponding data provide a limited basis for understanding voter behavior. If we are to improve our basis for recovering voter response information from partial-incomplete data we must introduce some structure into the modeling process. One bit of structure comes from the realization that the conditional probabilities β_{jk} must satisfy the row sum, $\sum_{k=1}^{c} \beta_{jk} = 1$, and column sum, $\sum_{j=1}^{g} \beta_{jk} N_{j\cdot} = N_{\cdot k}$, conditions. Some additional structure may be imposed based on a substantive theory about the particular behavior being examined and the elicitation of prior-nonsample information, and we can exploit this model structure to facilitate presentation and interpretation.

If we make use of the column sum conditions, we have the relationship

$$ n_{i\cdot k} = \sum_{j=1}^{g} n_{ij\cdot} \beta_{ijk} \; , \tag{1.1} $$

for $i = 1, \ldots, m$ and $k = 1, \ldots, c$. To formalize our notation, we let $\mathbf{x}(i) = (n_{i1\cdot} \; n_{i2\cdot} \; \cdots \; n_{ig\cdot})'$ represent the $(g \times 1)$ vector of proportions for each of the groups $j = 1, \ldots, g$ in precinct i, and let $\boldsymbol{y}(i) = (n_{i\cdot 1} \; n_{i\cdot 2} \; \cdots \; n_{i\cdot c})'$ represent the $(c \times 1)$ sample outcome vector of vote proportions for each candidate $k = 1, \ldots, c$ in precinct i. Then, the relationship among the observed marginal proportions and unknown conditional probabilities may be written as

$$ \boldsymbol{y}'(i) = \mathbf{x}'(i) \mathbf{B}(i) \; . \tag{1.2} $$

The component $\mathbf{B}(i) = (\boldsymbol{\beta}_{i1} \; \boldsymbol{\beta}_{i2} \; \cdots \; \boldsymbol{\beta}_{ic})$ is an unknown and unobservable $(g \times c)$ matrix of conditional probabilities and $\boldsymbol{\beta}_{ik} = (\beta_{i1k} \; \beta_{i2k} \; \cdots \; \beta_{igk})'$ is the $(g \times 1)$ vector of conditional probabilities associated with precinct i and candidate k. If we rewrite $\mathbf{B}(i)$ in $(gc \times 1)$ vectorized form as $\boldsymbol{\beta}(i) = \mathrm{vec}(\boldsymbol{\beta}(i)) = (\boldsymbol{\beta}'_{i1} \; \boldsymbol{\beta}'_{i2} \; \cdots \; \boldsymbol{\beta}'_{ic})'$, then we may rewrite (1.2) as

$$ \begin{bmatrix} y_1(i) \\ y_2(i) \\ \vdots \\ y_c(i) \end{bmatrix} = \begin{bmatrix} \mathbf{x}'(i) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}'(i) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}'(i) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{i1} \\ \boldsymbol{\beta}_{i2} \\ \vdots \\ \boldsymbol{\beta}_{ic} \end{bmatrix} \tag{1.3} $$

or more compactly as $\boldsymbol{y}(i) = (\mathbf{I}_c \otimes \mathbf{x}'(i)) \boldsymbol{\beta}(i) = \mathbf{X}(i) \boldsymbol{\beta}(i)$ where $\mathbf{X}(i) = (\mathbf{I}_c \otimes \mathbf{x}'(i))$ and $\otimes$ denotes the Kronecker product. We may extend the formulation to include observations for $m \geq 2$ precincts by stacking the $\boldsymbol{y}(i)$ and $\boldsymbol{\beta}(i)$

Table 1.1 *Known and Unknown Components of the Voter Problem*

Group	Candidate 1	2	3	4	Count
1	$\beta_{11}N_{1\cdot}$	$\beta_{12}N_{1\cdot}$	$\beta_{13}N_{1\cdot}$	$\beta_{14}N_{1\cdot}$	$N_{1\cdot}$
2	$\beta_{21}N_{2\cdot}$	$\beta_{22}N_{2\cdot}$	$\beta_{23}N_{2\cdot}$	$\beta_{24}N_{2\cdot}$	$N_{2\cdot}$
3	$\beta_{31}N_{3\cdot}$	$\beta_{32}N_{3\cdot}$	$\beta_{33}N_{3\cdot}$	$\beta_{34}N_{3\cdot}$	$N_{3\cdot}$
	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$N_{\cdot 4}$	N

vectors to form

$$
\begin{bmatrix} \boldsymbol{y}(1) \\ \boldsymbol{y}(2) \\ \vdots \\ \boldsymbol{y}(m) \end{bmatrix} = \begin{bmatrix} \mathbf{X}(1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}(2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}(m) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}(1) \\ \boldsymbol{\beta}(2) \\ \vdots \\ \boldsymbol{\beta}(m) \end{bmatrix} \tag{1.4}
$$

or $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta}$.

Given relationships (1.2) to (1.4) as a way of modeling the underlying data process, we view each election as an experiment. The sample data underlying Table 1.1 is viewed as the outcome of an election-experiment. Consequently, we represent these sample data as having a systematic component (1.2) and a random component ε_{ik} and write the statistical model expressing the data sampling process as

$$
n_{i\cdot k} = \sum_{j=1}^{g} n_{ij\cdot}\beta_{ijk} + \varepsilon_{ik} \tag{1.5}
$$

or

$$
\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} , \tag{1.6}
$$

where the noise vector $\boldsymbol{\epsilon}$ is supported on a non-empty and bounded set and is assumed to have mean $\mathrm{E}\left[\boldsymbol{\epsilon}\right] = \mathbf{0}$, and finite covariance matrix $\boldsymbol{\Sigma}$. The errors represent sampling variation in the observed column sums (n_{ik}) relative to the true but unobserved marginal probability that voters in precinct i vote for candidate k. Using this weak model specification, we initially solve the problem by using the observed data outcomes $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta}$ to represent the population moments, $\mathrm{E}\left[\boldsymbol{y}\right] = \mathrm{E}\left[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\right]$. Under this form, the absence of sampling errors and other stochastic noise components in (1.2)–(1.4) implies that the problem of recovering $\boldsymbol{\beta}$ from observed $\boldsymbol{y}$ and $\mathbf{X}$ is a pure inverse problem. For each precinct-specific problem (1.3), note that the matrix $\mathbf{X}(i)$ has dimension $(c \times gc)$ and is underdetermined and generally not invertible. Thus, under traditional mathematical inversion procedures, the voter pure inverse (VPI) problem is said to be ill-posed, and

the solution space for the problem contains arbitrary parameters. The question we now face is whether or not there is a plausible basis for reasoning in situations like this where the information we possess specifies only a feasible set of functions. Moving in the direction of a plausible solution basis, we note that the unknown conditional probabilities must satisfy some additional conditions such as additivity and non-negativity, and the solution to the pure inverse problem (1.2) must satisfy the estimating functions $y = \mathbf{X}\beta$.

1.3 Some Information Theoretic Solutions

1.3.1 Choosing the Criterion Function

Given the inverse model specified in Section 1.2.1, the long journey in defining a solution begins with the selection of a goodness-of-fit criterion. If we recognize and maintain the distinction that the unknown elements β_{ijk} are conditional probabilities rather than joint probabilities, then our pure voting inverse model is similar to allocating values to each of the cells in a contingency table. Consequently, the Cressie-Read power-divergence (CRPD) statistic (Cressie and Read, 1984; Read and Cressie, 1988; Baggerly, 1998) is a pseudo-distance measure that may be used to compare elements in the set of feasible conditional probabilities implied by the available data. For a discrete probability distribution $\mathbf{w}$ defined with respect to $i = 1, \ldots, n$ possible outcomes, the CRPD statistic

$$I(\mathbf{w}, \boldsymbol{q}, \lambda) = \frac{2}{\lambda(1 + \lambda)} \sum_i w_i \left[\left(\frac{w_i}{q_i} \right)^{\lambda} - 1 \right] , \tag{1.7}$$

measures the pseudo-distance between $\mathbf{w}$ (i.e., conditional probabilities in the VPI problem) and a set of reference weights $\boldsymbol{q}$. The reference weights may be based on additional or prior information that the researcher may want to bring to bear upon the estimation. The discrete weights must satisfy $(w_i, q_i) \in (0, 1) \times (0, 1) \ \forall \ i$ and $\sum_i w_i = \sum_i q_i = 1$. Read and Cressie note that (1.7) encompasses a family of empirical likelihood estimation objective functions that includes

1. Kullback-Leibler directed divergence or discrimination information statistic (Kullback, 1959; Gokhale and Kullback, 1978)

$$I(\mathbf{w}, \boldsymbol{q}, \lambda \to 0) \propto \sum_{i=1}^{n} w_i \ln \left(\frac{w_i}{q_i} \right) \tag{1.8}$$

and

$$I(\mathbf{w}, \boldsymbol{q}, \lambda \to -1) \propto \sum_{i=1}^{n} q_i \ln \left(\frac{q_i}{w_i} \right) \tag{1.9}$$

Note that $I(\mathbf{w}, \boldsymbol{q}, \lambda \to 0) + I(\mathbf{w}, \boldsymbol{q}, \lambda \to -1)$ is a symmetric distance function.

2. Pearson's chi-square statistic (Pearson, 1900)

$$I(\mathbf{w}, \boldsymbol{q}, \lambda = 1) = \sum_{i=1}^{n} \frac{(w_i - q_i)^2}{q_i} \qquad (1.10)$$

3. Modified chi-square statistic (Neyman, 1949)

$$I(\mathbf{w}, \boldsymbol{q}, \lambda = -2) = \sum_{i=1}^{n} \frac{(q_i - w_i)^2}{w_i} \qquad (1.11)$$

4. Squared Matusita or Hellinger distance

$$I(\mathbf{w}, \boldsymbol{q}, \lambda = -1/2) \propto \sum_{i=1}^{n} \left(\sqrt{w_i} - \sqrt{q_i}\right)^2 \qquad (1.12)$$

Read and Cressie note that the CRPD statistic is strictly convex in its arguments and may be used as a criterion function for minimum distance estimation. Given uniform reference weights $q_i = n^{-1}$, $\forall\, i$, the negative of (1.7) also encompasses other prominent statistics:

1. Empirical likelihood statistic (Owen, 1988, 1990)

$$-I(\mathbf{w}, \boldsymbol{q}, \lambda \to -1) \propto \sum_{i=1}^{n} \ln(w_i) \qquad (1.13)$$

2. Shannon's entropy (Shannon, 1948) or exponential empirical likelihood (Di Cicco and Romano, 1999; Corcoran, 2000) statistic

$$-I(\mathbf{w}, \boldsymbol{q}, \lambda \to 0) \propto -\sum_{i=1}^{n} w_i \ln(w_i) \qquad (1.14)$$

3. Simpson or Gini statistic

$$-I(\mathbf{w}, \boldsymbol{q}, \lambda = 1) = 1 - \sum_{i=1}^{n} w_i^2 \qquad (1.15)$$

In these cases, the minimum distance estimation problem is solved by maximizing the criterion function with respect to $\mathbf{w}$.

1.3.2 Solution to a Pure Inverse Problem

Under the minimum CRPD estimation criterion, an estimator for the VPI problem may be derived by minimizing the estimation criterion (1.7) for this problem

$$\frac{2}{\lambda(1+\lambda)} \sum_{i=1}^{m} \sum_{j=1}^{g} \sum_{k=1}^{c} \beta_{ijk} \left[\left(\frac{\beta_{ijk}}{q_{ijk}}\right)^{\lambda} - 1 \right], \qquad (1.16)$$

(given some λ) subject to the column-sum condition (1.1) and the row-sum (additivity) condition

$$\sum_{k=1}^{c} \beta_{ijk} = 1 \qquad \forall\, i, j \, . \tag{1.17}$$

The Lagrange expression for this constrained minimization problem is

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \;=\;& \frac{2}{\lambda(1+\lambda)} \sum_{i=1}^{m} \sum_{j=1}^{g} \sum_{k=1}^{c} \beta_{ijk} \left[\left(\frac{\beta_{ijk}}{q_{ijk}} \right)^{\lambda} - 1 \right] \\
&+ \sum_{i=1}^{m} \sum_{k=1}^{c} \alpha_{ik} \left(n_{i\cdot k} - \sum_{j=1}^{g} n_{ij\cdot}\beta_{ijk} \right) \\
&+ \sum_{i=1}^{m} \sum_{j=1}^{g} \gamma_{ij} \left(1 - \sum_{k=1}^{c} \beta_{ijk} \right) \, ,
\end{aligned} \tag{1.18}
$$

where α_{ik} and γ_{ij} are Lagrange multipliers for constraints (1.1) and (1.17), respectively. The necessary condition for $\widehat{\beta}_{ijk}$ is

$$\frac{\partial L}{\partial \beta_{ijk}} = \frac{2}{\lambda} \left(\frac{\widehat{\beta}_{ijk}}{q_{ijk}} \right)^{\lambda} - \frac{2}{\lambda(1+\lambda)} - \widehat{\alpha}_{ik}\, n_{ij\cdot} - \widehat{\gamma}_{ij} = 0 \, , \tag{1.19}$$

and the solution for the conditional probabilities is

$$\widehat{\beta}_{ijk} = q_{ijk} \left[\frac{1}{1+\lambda} + \frac{\lambda}{2} \left(\widehat{\alpha}_{ik}\, n_{ij\cdot} + \widehat{\gamma}_{ij} \right) \right]^{1/\lambda} \, . \tag{1.20}$$

In general, the solution does not have a closed-form expression and must be stated in intermediate form as a function of the optimal Lagrange multipliers, $\widehat{\alpha}_{ik}$ and $\widehat{\gamma}_{ij}$. Consequently, the optimal values of the Lagrange multipliers must be numerically determined. We note that as $\lambda \to 0$ in (1.16), the estimating criterion is

$$\sum_{i=1}^{m} \sum_{j=1}^{g} \sum_{k=1}^{c} \beta_{ijk} \, \ln \left(\frac{\beta_{ijk}}{q_{ijk}} \right) \, , \tag{1.21}$$

and the intermediate solution for the constrained optimal β_{ijk} is

$$\widehat{\beta}_{ijk} = \frac{q_{ijk}\, \exp \left(\widehat{\alpha}_{ik}\, n_{ij\cdot} \right)}{\sum_{k=1}^{c} q_{ijk}\, \exp \left(\widehat{\alpha}_{ik}\, n_{ij\cdot} \right)} \, . \tag{1.22}$$

The elements $\widehat{\alpha}_{ik}$ are the optimal values of the Lagrange multipliers on constraint (1.4). Under uniform reference weights ($q_{ijk} = c^{-1}, \; \forall\, i, j$), the negative criterion is proportional to

$$- \sum_{i=1}^{m} \sum_{j=1}^{g} \sum_{k=1}^{c} \beta_{ijk} \, \ln \left(\beta_{ijk} \right) \, , \tag{1.23}$$

and the minimum CRPD problem is known in the information theory literature as Jaynes' method of maximum entropy (Jaynes, 1957a, 1957b) for pure inverse problems. Illustrative examples using real and synthetic data are given in Appendix A.

1.3.3 Incorporating Bounds on the Conditional Probabilities

Given the high degree of ambiguity resulting from the aggregate data, we can follow Duncan and Davis (1953) and use (1.1) to refine the constraint set on the conditional probabilities by placing upper and lower bounds on each β_{ijk}. As indicated by King (1997), the constraint (1.1) implies that the lower bound on β_{ijk} is $Z_{ijk1} = \max\left(0, \left(n_{i\cdot k} - \left(1 - n_{ij\cdot}\right)\right)/n_{ij\cdot}\right)$, and the upper bound is $Z_{ijk2} = \min\left(1, n_{i\cdot k}/n_{ij\cdot}\right)$. Given the bounds, β_{ijk} may be expressed as a convex combination, $\beta_{ijk} = \sum_{h=1}^{2} \varphi_{ijkh} Z_{ijkh}$ for $\varphi_{ijkh} \geq 0$ such that $\varphi_{ijk1} + \varphi_{ijk2} = 1$. In this case, we may specify reference weights q_{ijkh} on each of the upper and lower bounds such that $\beta_{ijk}^{0} = \sum_{h=1}^{2} q_{ijkh} Z_{ijkh}$ is a pre-sample estimate of the unknown β_{ijk}.

After incorporating the bounding information, the reformulated VPI problem may now be solved by minimizing

$$\frac{2}{\lambda(1+\lambda)} \sum_{i=1}^{m} \sum_{j=1}^{g} \sum_{k=1}^{c} \sum_{h=1}^{2} \varphi_{ijkh} \left[\left(\frac{\varphi_{ijkh}}{q_{ijkh}} \right)^{\lambda} - 1 \right] , \qquad (1.24)$$

subject to reparameterized versions of (1.1) and (1.17)

$$n_{i\cdot k} = \sum_{j=1}^{g} n_{ij\cdot} \sum_{h=1}^{2} \varphi_{ijkh} Z_{ijkh} , \qquad (1.25)$$

$$1 = \sum_{k=1}^{c} \sum_{h=1}^{2} \varphi_{ijkh} Z_{ijkh} , \qquad (1.26)$$

plus the additivity constraint on the new weights

$$\varphi_{ijk1} + \varphi_{ijk2} = 1 . \qquad (1.27)$$

Setting up and solving the first order conditions leads to the solution

$$\widehat{\varphi}_{ijkh} = q_{ijkh} \left[\frac{1}{1+\lambda} + \frac{\lambda}{2} \left(\widehat{\alpha}_{ik} n_{ij\cdot} Z_{ijkh} + \widehat{\gamma}_{ij} Z_{ijkh} + \rho_{ijk} \right) \right]^{1/\lambda} , \qquad (1.28)$$

where the point estimator of the bounded conditional probability is

$$\widehat{\beta}_{ijk} = \sum_{h=1}^{2} \widehat{\varphi}_{ijkh} Z_{ijkh} . \qquad (1.29)$$

If we let $\lambda \to 0$ in the criterion function (1.16), we are led to the constrained minimization problem

$$\sum_{i=1}^{m}\sum_{j=1}^{g}\sum_{k=1}^{c}\sum_{h=1}^{2} \varphi_{ijkh} \, \ln \left(\frac{\varphi_{ijkh}}{q_{ijkh}} \right) \, , \tag{1.30}$$

subject to the constraints (1.25) to (1.27). The intermediate solution for the constrained optimal φ_{ijkh} may be expressed as

$$\widehat{\varphi}_{ijkh} = \frac{q_{ijkh} \, \exp \left(\widehat{\alpha}_{ik} n_{ij\cdot} Z_{ijkh} + \widehat{\gamma}_{ij} Z_{ijkh} \right)}{\sum_{h=1}^{2} q_{ijkh} \, \exp \left(\widehat{\alpha}_{ik} n_{ij\cdot} Z_{ijkh} + \widehat{\gamma}_{ij} Z_{ijkh} \right)} \, , \tag{1.31}$$

and the estimator of the conditional probabilities is (1.29). Illustrative examples for these formulations using real and synthetic data are given in Appendix A.

1.3.4 The Noisy Voter Inverse Problem

The pure voter inverse problem of Sections 1.3.2 and 1.3.3 is one plausible way to model information recovery from aggregate election data. However, if we view each election as an experiment, then Table 1.1 represents the outcome of an election-experiment. Because it may be unrealistic to assume that the vote counts and the shares (proportions) are perfectly observed, in the spirit of much of the research in the ecological inference area, we turn to the following sampling model that has both systematic and stochastic components that represent the sampling process as in (1.5)

$$n_{i\cdot k} = \sum_{j=1}^{g} n_{ij\cdot} \beta_{ijk} + \varepsilon_{ik} \, , \tag{1.32}$$

or in the form of a linear statistical model

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \, . \tag{1.33}$$

The random $(mgc \times 1)$ noise vector $\boldsymbol{\epsilon}$ is assumed to have mean $\mathrm{E}\left[\boldsymbol{\epsilon}\right] = \mathbf{0}$ and finite covariance matrix $\boldsymbol{\Sigma}$. At this point, we assume that the $\mathbf{X}$'s are measured without error. We refer to the resulting estimation problem as the voter noisy inverse (VNI) problem.

1.3.4.1 Incorporating Bounds on the Characteristics of the Noise

The properties of $\varepsilon_{ik} \in [0, 1]$ may be derived from the known properties of $n_{i\cdot k}$ in (1.5) or (1.32). First, we may be able to refine the plausible subset of the error space for a given sample by using the method of bounds to determine the plausible upper and lower bounds on ε_{ik}. The largest possible positive difference between $n_{i\cdot k}$ and the systematic component occurs if the conditional probabilities assigned to each group j in column k are zero (i.e., $\beta_{ijk} = 0$), and the upper bound is

simply $n_{i \cdot k}$. Accordingly, the largest possible negative difference occurs if $\beta_{ijk} = 1$ for each j in column k, and the lower bound is $n_{i \cdot k} - 1$.

Second, we may refine the bounds to reflect the statistical properties of ϵ. Under the standard sampling conditions commonly assumed for the problem, ϵ is a mean zero process with finite covariance Σ. Further, $n_{i \cdot k}$ is a $\sqrt{N_i}$-consistent estimator of the marginal probability $\beta_{i \cdot k}$ such that $n_{i \cdot k} \overset{p}{\to} \beta_{i \cdot k}$ and is asymptotically normal as $\sqrt{N_i} \left(n_{i \cdot k} - \beta_{i \cdot k} \right) \overset{d}{\to} \mathrm{N} \left(0, \beta_{i \cdot k} \left(1 - \beta_{i \cdot k} \right) \right)$. Consequently, we know the bounds should be centered about zero. Let $V_{ik1} = \delta_{ik} \left(n_{i \cdot k} - 1 \right) / \sqrt{N_i}$ be the lower bound and $V_{ik2} = \delta_{ik} n_{i \cdot k} / \sqrt{N_i}$ be the upper bound for each error term (where $\delta_{ik} > 0$ may be distinct for each i and k). To directly impose the mean-zero property of ϵ, we may specify symmetric (about zero) bounds, $V_{ik1} = -\delta_{ik} \max \left(n_{i \cdot k}, 1 - n_{i \cdot k} \right) / \sqrt{N_i}$ and $V_{ik2} = -V_{ik1}$.

Given the error bounds, each ε_{ik} may be expressed as a convex combination

$$\varepsilon_{ik} = w_{ik1} V_{ik1} + w_{ik2} V_{ik2} , \tag{1.34}$$

where $w_{ikr} > 0$ for $r = 1, 2$ and $w_{ik1} + w_{ik2} = 1$. Then, (1.32) may be reformulated as

$$n_{i \cdot k} = \sum_{j=1}^{g} n_{ij \cdot} \sum_{h=1}^{2} Z_{ijkh} \varphi_{ijkh} + \sum_{r=1}^{2} V_{ikr} w_{ikr} . \tag{1.35}$$

By construction, there exist simplex-valued weights $\{\varphi_{ijkh}\}$ and $\{w_{ikr}\}$ such that (1.35) holds for the observed sample. Through (1.35), the VNI problem may now be based on the linear statistical model (1.33) plus the bounding information on β_{ijk} and ε_{ik}. Thus, the VNI problem may be solved by formulating it as a minimum distance estimation problem and determining an appropriate set of weights for the unknown conditional probabilities and error components. As before, the problem specification allows for reference weights on the unknown parameters, q_{ijkh} for φ_{ijkh} and q_{ikr}^{w} for w_{ikr}.

1.3.4.2 The Solution

As in Sections 1.3.2 and 1.3.3, we solve the extended VNI problem by minimizing the CRPD criterion subject to the complete set of constraints. In particular, we minimize

$$\sum_{i=1}^{m} \sum_{j=1}^{g} \sum_{k=1}^{c} \sum_{h=1}^{2} \varphi_{ijkh} \ln \left(\frac{\varphi_{ijkh}}{q_{ijkh}} \right) + \sum_{i=1}^{m} \sum_{k=1}^{c} \sum_{r=1}^{2} w_{ikr} \ln \left(\frac{w_{ikr}}{q_{ikr}^{w}} \right) , \tag{1.36}$$

subject to (1.35) plus

$$\sum_{k=1}^{c} \sum_{h=1}^{2} \varphi_{ijkh} Z_{ijkh} = 1 , \tag{1.37}$$

$$\varphi_{ijk1} + \varphi_{ijk2} \;=\; 1 \,, \tag{1.38}$$

$$w_{ik1} + w_{ik2} \;=\; 1 \,. \tag{1.39}$$

The necessary conditions yield the intermediate solutions for the weights

$$\widehat{\varphi}_{ijkh} \;=\; \frac{q_{ijkh} \exp\left(\widehat{\alpha}_{ik}\, n_{ij\cdot}\, Z_{ijkh} + \widehat{\gamma}_{ij} Z_{ijkh}\right)}{\sum_{h=1}^{2} q_{ijkh} \exp\left(\widehat{\alpha}_{ik}\, n_{ij\cdot}\, Z_{ijkh} + \widehat{\gamma}_{ij} Z_{ijkh}\right)} \,, \tag{1.40}$$

$$\widehat{w}_{ikr} \;=\; \frac{q_{ikr}^{w} \exp\left(\widehat{\alpha}_{ik} V_{ikr}\right)}{q_{ik1}^{w} \exp\left(\widehat{\alpha}_{ik} V_{ik1}\right) + q_{ik2}^{w} \exp\left(\widehat{\alpha}_{ik} V_{ik2}\right)} \,. \tag{1.41}$$

After the optimal Lagrange multipliers $\widehat{\alpha}_{ik}$ and $\widehat{\gamma}_{ij}$ are numerically determined, the estimates $\widehat{\beta}_{ijk}$ are computed as in (1.29). In Appendix A, a data set is used to illustrate the noisy inverse formulation and compare it to the pure inverse formulation. In general, the introduction of the noise components weakens the constraints and moves the estimate in the direction of the initial unbounded pure outcome.

The ill-posed and underdetermined character of the pure and noisy inverse problems implies that a unique solution does not exist. Each "solution" is merely an algorithm for inferring a function (conditional probabilities) that is consistent with the available information (constraints) and the estimation criterion. Accordingly, we have developed a feasible solution method based on information theoretic-empirical likelihood concepts and tools. Under the noise inverse formulation, it is possible to demonstrate that under standard regularity conditions on $n_{i\cdot k}$ (stated in Section 1.3.4.1) that $\hat{\beta}_{jk}$ is a $\sqrt{N_i}$–consistent and asymptotically normal estimator. The large-sample properties of the estimator presents a basis for inference about the minimum CRPD model of voting behavior.

1.3.5 Remarks

One of the great queries of political science revolves around the ideal of representation. Do our political institutions promote or inhibit fair representation of the masses? This is a difficult question to answer, and there are few mechanisms through which we can gain insight into this query. One mechanism, however, is the election process. Elections can be seen as natural experiments where we are able to observe repeatedly the behavior of the citizenry and its response to political institutions over time. Although it is difficult to make large-scale changes in our political structures, policies and platforms certainly change in response to each election, and this process is ongoing. To gain maximal insight, we would ideally like to know how preferences map to choices in a variety of contextual settings. Surveys usually cannot capture these varieties of settings, whereas aggregate returns can. Moreover, an analysis of aggregate data allows us to study those more rare instances in which institutions do shift and we want to see how that change in structure alters the mapping of preferences to choice. The formulations in this section allow us to tap into these types of questions on a macro level.

We can observe how a macro unit such as a precinct behaves across elections and through time. We can, moreover, through the information theoretic formulations, observe how this behavior changes as a function of precinct characteristics such as urban/rural or minority composition or the strengths of major party affiliations.

These formulations are based on aggregate data, and so the results necessarily apply directly to the aggregate units only. The connection to individual-level behavior is clearly indirect. Nonetheless, if one needs to provide an interpretation of individual behavior based on aggregate data, the estimated conditional probabilities from the information theoretic approach is a plausible basis. These conditional probability estimates are admittedly only one way to summarize the aggregate data, but using this trajectory to arrive at a "solution" is attractive in several senses. First, the information theoretic procedure provides a solution to the ill-posed inverse problem that is consistent with the possible underlying data generating process. Second, this procedure provides an especially appealing solution because the outcomes represent voter counts that could have occurred in the greatest number of ways given the data constraints (see Section 2.2 in Golan, Judge, and Miller for details). Third, political science theories rarely provide an adequate basis for specifying the random mechanism by which the observed data are generated. As previously noted, the proposed approach is semiparametric and does not require a fully specified likelihood function. Fourth, the information theoretic procedures allow one to stay within the general framework while using additional non-sample information to condition the solution.

1.4 Recovering Information on Individual Behavior

In the approach to information recovery for the inverse problems in Section 1.3, we only use the observed macro data relating to voter groups and candidates. Since one uses aggregate data as an input, one gets information relevant at the aggregate level as output. For many voter behavior questions, these aggregated estimates may not provide an adequate basis for inference in either a positive or normative sense. Ultimately, we are interested in micro-individual voter behavior, and this is the topic to which we now turn.

Viewing each election as an experiment, we use (1.32) and (1.33) as a basis for modeling the sampling process. We noted that it may be unrealistic to assume that the vote counts and voter group shares (i.e., the n_{ij}'s or the $\mathbf{x}$'s) are measured without error. If we let $\mathbf{x}^*$ be the observed voter group shares from (1.33) and $\mathbf{x}$ be the true unobservable voter shares, then we may model $\mathbf{x}^*$ as

$$\mathbf{x}^* = \mathbf{x} + \mathbf{u}, \tag{1.42}$$

where $\mathbf{u}$ is an unobserved noise vector. Therefore, the underlying statistical model is

$$y = \mathbf{x}\beta + \epsilon, \tag{1.43}$$

but the observable version of (1.33) is

$$y = \mathbf{x}^*\beta + \epsilon^* \,, \tag{1.44}$$

where

$$\epsilon^* = \epsilon - \mathbf{u}\beta \,. \tag{1.45}$$

If the measurement errors in $\mathbf{x}$ are independent of ϵ and mutually uncorrelated, then ϵ^* is a mean zero noise vector with covariance $\boldsymbol{\Sigma} + \beta\beta'\sigma_u^2\mathbf{I} = \boldsymbol{\Omega}$. In the statistical model based on the observable data, $\mathbf{x}^*$ is correlated with the random matrix ϵ^*. Thus, the usual linear model condition that the right-hand-side explanatory variables are orthogonal in expectation to the error process is violated. Further, traditional estimation rules based on $\mathrm{E}\left[\mathbf{x}^{*\prime}\epsilon^*\right] = \mathbf{0}$ will have questionable statistical properties when this condition does not hold.

To mitigate the impact of the measurement errors, we use additional information that we identify in the form of instrumental variables. This source of information makes use of the fact that precincts may vary in terms of their individual demographic characteristics and that this variation may be related to the corresponding unknown and unobservable voter group shares and conditional probabilities. Given the economic, political, and social differences between precincts, it seems likely that the group shares and conditional probabilities β_{ijk} may vary over individuals, precincts, and/or time. To reflect this potential heterogeneity in the micro behavior, we assume that the β_{ijk}'s are conditional on a set of explanatory-instrumental variables, and that these covariates reflect the individual, spatial, or temporal differences in voter decisions. As such, the instrumental variable (IV) approach provides a method for estimating causal effects in a measurement error or simultaneous equation model framework. The covariates may include measures of economic performance such as the local level of unemployment, political characteristics such as incumbency, or demographic variables such as average age of the electorate. Using this information, along with the observed macro data discussed in Section 1.3, it is possible to form a set of estimating equations as a basis for recovering the unknown conditional probabilities and identifying the impact of the explanatory variables on the corresponding conditional probabilities. The ultimate success of the moment-based specification depends on a plausible theory of micro voter behavior that helps to identify the important behavior conditioning factors.

Returning to the statistical model (1.42) where the observed $\mathbf{x}^*$'s are now stochastic explanatory variables that are correlated with the noise vector ϵ^*, one useful way to model the data sampling process is to consider y and $\mathbf{x}^*$ as endogenous-jointly determined random variables. In this context, the statistical model becomes a system of relations

$$y = \mathbf{x}^*\beta + \epsilon^* \,, \tag{1.46}$$

and

$$\mathbf{x}^* = \mathbf{A}\pi + \mathbf{u} \,, \tag{1.47}$$

where $\mathbf{A}$ is a set of instrumental variables that are correlated with $\mathbf{x}^*$ but uncorrelated with $\boldsymbol{\epsilon}^*$ and $\mathbf{u}$. Under this formulation, the simultaneous or structural equation statistical model results, and traditional estimation and inference procedures apply directly to the model (see Chapter 17 in Mittelhammer, Judge, and Miller for more details).

In practice, the source of measurement error is varied and specific to the application being considered. In general, we expect that some measurement error will be evident in the $\mathbf{X}$ variable. For example, in the Voting Rights context, errors in the $\mathbf{X}$ variable are commonplace, since the variable of interest, racial turnout proportions, is rarely attainable. Instead, one must reply on a proxy variable such as racial registration proportions or racial population proportions. Using a proxy variable leads to the type of measurement error mentioned above.

1.4.1 Moment-Based Model Formulation

To link the β_{ijk} to the explanatory-instrumental variables, we rewrite the noisy inverse statistical model (1.32) as

$$\mathbf{A}'\boldsymbol{y} = \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \mathbf{A}'\boldsymbol{\epsilon} \,. \tag{1.48}$$

The explanatory variables $\mathbf{A}$ in (1.48) are assumed to be uncorrelated with the noise components. Consequently, we can form the following set of estimating equations

$$\mathrm{E}\left[\mathbf{A}'\left(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}\right)\right] = \mathbf{0} \,, \tag{1.49}$$

and the sample analog

$$T^{-1}\mathbf{A}'\left(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}\right) \overset{p}{\to} \mathbf{0} \ \text{ as } \ T \to \infty \,. \tag{1.50}$$

The individual components of the moment conditions may be stated in scalar form as

$$T^{-1}\sum_{t=1}^{T} \mathbf{A}'_{ti}\left[n_{ti\cdot k} - \sum_{j=1}^{g} n_{tij\cdot}\beta_{tijk}\right] = \mathbf{0} \,, \tag{1.51}$$

for each i and k. To allow for heteroskedasticity across precincts and possible temporal correlation, we assume regularity conditions on $\boldsymbol{\epsilon}$ such that (1.51) holds under an appropriate weak law.

The moment condition (1.51) may be extended to include the reparameterized conditional probabilities β_{tijk} and noise components ε_{tik} as in Section 1.3. The voter inverse problem with noise and time-varying conditional probabilities may be solved by minimizing the CRPD objective function

$$I\left(\varphi, \mathbf{w}, \lambda\right) \;=\; \frac{2}{\lambda(1+\lambda)}\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{g}\sum_{k=1}^{c}\sum_{h=1}^{2}\varphi_{tijkh}\left[\left(\frac{\varphi_{tijkh}}{q_{tijkh}}\right)^{\lambda} - 1\right]$$

$$+ \frac{2}{\lambda(1+\lambda)} \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{k=1}^{c} \sum_{r=1}^{2} w_{tikr} \left[\left(\frac{w_{tikr}}{u_{tikr}} \right)^{\lambda} - 1 \right], (1.52)$$

subject to the estimating equations

$$\sum_{t=1}^{T} \mathbf{A}'_{ti} \left[n_{ti\cdot k} - \sum_{j=1}^{g} n_{tij\cdot} \sum_{h=1}^{2} Z_{tijkh} \varphi_{tijkh} - \sum_{r=1}^{2} V_{tikr} w_{tikr} \right] = \mathbf{0}, \quad (1.53)$$

plus the additivity conditions

$$\sum_{k=1}^{c} \varphi_{tijkh} Z_{tijkh} = 1, \tag{1.54}$$

$$\varphi_{tijk1} + \varphi_{tijk2} = 1, \tag{1.55}$$

$$w_{tik1} + w_{tik2} = 1. \tag{1.56}$$

The intermediate solution may be stated in terms of the Lagrange multipliers

$$\widehat{\varphi}_{tijkh} = \frac{q_{tijkh} \exp\left(\widehat{\boldsymbol{\alpha}}'_{ik} \mathbf{A}_{ti} Z_{tijkh} n_{tij\cdot} \right)}{\sum_{h=1}^{2} q_{tijkh} \exp\left(\widehat{\boldsymbol{\alpha}}'_{ik} \mathbf{A}_{ti} Z_{tijkh} n_{tij\cdot} \right)}. \tag{1.57}$$

The minimum CRPD estimator of the time-varying conditional probability is

$$\widehat{\beta}_{tijk} = \sum_{h=1}^{2} \widehat{\varphi}_{tijkh} Z_{tijkh}. \tag{1.58}$$

The Lagrange multiplier $\widehat{\alpha}_{ik}$ provides a basis for evaluating the impact of the instrumental variables on the solution.

In general, the estimators for the moment-based model formulation will also be consistent and asymptotically normal under standard regularity conditions. For example, the consistency result stated in Equation (1.50) and a related assumption regarding the asymptotic normality of $T^{-1/2} \mathbf{A}'(\mathbf{Y} - \mathbf{X}\beta)$ may be used to establish the asymptotic properties. To illustrate the basic statistical properties of the moment-based formulation, we conduct a series of Monte Carlo simulation exercises for a cross-sectional version of the model with $m = 20$ and $m = 50$ precincts or districts. Overall, the replicated estimates of the model parameters exhibit smaller sample bias and variance as the number of precincts m increases. We discuss further details regarding the composition of the replicated sampling process and the simulation results in the Appendix.

1.4.1.1 Remarks

Applications of the ecological inference problem are often in areas where the estimates are highly consequential. For instance, in the Voting Rights arena, the decision of a judge to grant or deny relief under the Voting Rights Act turns entirely

on an ecological inference. How the system of representation plays out in our democracy is closely tied to how this type of legislation is enforced. Hence, not accounting for measurement error in this context especially could have great ramifications. Moreover, this is a circumstance where measurement error is known to pose a problem. In particular, the voter group shares are often based on registration rates (which can be attained for a small set of localities) or populations figures (which are easily attainable), but the variable of interest is racial turnout rates (which are very difficult to obtain). Using one as a proxy for the other may be necessary, but also clearly problematic.

Some scholars have suggested a "double regression" approach to alleviate this problem (Kousser 1973; Grofman, Handley, and Niemi 1992). This method sports the same idea as the minimum CRPD method, but does not take any additional information into account. The proposed instrumental variables approach allows one to incorporate the large literature on voter turnout to help mitigate the impact of the undisputed measurement error. Although the success of this formulation is dependent on a plausible theory of micro-level behavior, the uncertainty can be assuaged by the reliance on solid empirical studies in an extensive substantive literature. While the formulations proposed here are at the macro level, they incorporate information that has been empirically verified at the micro level.

1.4.2 The Discrete Choice Voter Response Model

In this section, we focus on obtaining and using micro data about individual voters in a precinct. Our objective is to use this micro data to estimate the impact of political, social economic, and demographic variables on voter behavior and to recover the corresponding marginal (conditional) probabilities. We envision a situation where detailed survey data are collected on variables that characterize the voters in the precinct and indicate how each person voted in a particular contest. Given micro data that reflects the individual characteristics of a sample of voters, we model voter response as a discrete binary choice problem.

To develop this model, we use the unordered multinomial statistical response model. In this context, consider an unordered multinomial discrete choice problem with an experiment (survey) consisting of N trials (voters in a precinct), binary random variables $y_{1j}, y_{2j}, \ldots, y_{Nj}$ are observed. The binary outcomes $\{y_{ij}\}$ are observed for voters $i = 1, \ldots, N$ and candidates $j = 1, 2, \ldots, J$ in a given precinct. The candidate indices may be reordered without loss of generality such that the candidates represent J unordered categories. The observed outcome is $y_{ij} = 1$ if and only if voter i casts a vote for candidate j, and $y_{ij} = 0$ otherwise.

Let the probability that voter i casts a vote for candidate j be p_{ij} and assume that the voting decision is related to a set of explanatory variables $\mathbf{a}_i$ through the model

$$p_{ij}(\boldsymbol{\beta}) \equiv \mathrm{P}\left(y_{ij} = 1 \mid \mathbf{a}_i, \boldsymbol{\beta}_j\right) = G\left(\mathbf{a}_i'\boldsymbol{\beta}_j\right) > 0, \qquad (1.59)$$

for each i and j. In particular, $\mathbf{a}_i' = (a_{i1}, a_{i2}, \ldots, a_{iK})$, $\boldsymbol{\beta}_j$ is a $(K \times 1)$ vector of unknown response parameters, and $G(\cdot)$ is a function that links the probabilities p_{ij} with the linear combination $\mathbf{a}_i'\boldsymbol{\beta}_j$ such that $G(\mathbf{a}_i'\boldsymbol{\beta}_j) \in [0, 1]$ and $\sum_{j=1}^{J} G(\mathbf{a}_i'\boldsymbol{\beta}_j) = 1$.

Suppose the observed outcomes of y_{ij} are noisy such that the underlying binary random variables may be modeled as

$$Y_{ij} = G(\mathbf{a}_i'\boldsymbol{\beta}_j) + \varepsilon_{ij} = p_{ij} + \varepsilon_{ij} \,, \tag{1.60}$$

where the ε_{ij} are noise components. The binary response model may be written in matrix form as

$$\boldsymbol{Y} = \boldsymbol{p} + \boldsymbol{\epsilon} \,, \tag{1.61}$$

where each component is an $(NJ \times 1)$ vector. We assume $\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and that $\mathbf{cov}(\boldsymbol{\epsilon})$ is a finite positive semidefinite matrix. Note that this matrix is rank-deficient due to the additivity property of the choice probabilities, $\sum_{j=1}^{J} p_{ij} = 1$.

If we follow McFadden (1974), Manski and McFadden (1982), or Maddala (1983), we may solve the problem with the traditional maximum likelihood approach. Under the log-likelihood function

$$\ln \mathrm{L}(\boldsymbol{\beta}; \mathbf{a}) = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \ln \left(G(\mathbf{a}_i'\boldsymbol{\beta}_j) \right) \,, \tag{1.62}$$

the solution is the multinomial logit estimator if G is the logistic CDF and the multinomial probit estimator if G is the multivariate normal (Gaussian) CDF.

Rather than adopt a fully parametric specification, we extend the ideas outlined in Section 1.3 and use a moment-based approach for estimation and inference. In this context, we use the observed outcomes of $\boldsymbol{y}$ and the $(N \times K)$ matrix of explanatory variables $\mathbf{a}$ to recover information about the *unknown* and *unobservable* model components $\boldsymbol{p}$ and $\boldsymbol{\beta}$. For the multinomial choice problem, this information may be written as an inverse problem with noise that is linear in $\boldsymbol{p}$

$$(\mathbf{I}_J \otimes \mathbf{a}') \, \boldsymbol{y} = (\mathbf{I}_J \otimes \mathbf{a}') \, \boldsymbol{p} + (\mathbf{I}_J \otimes \mathbf{a}') \, \boldsymbol{\epsilon} \,. \tag{1.63}$$

The inverse problem has KJ moment relations and NJ unknown conditional probabilities. Assuming the orthogonality condition $\mathrm{E}\left[(\mathbf{I}_J \otimes \mathbf{a}') \, \boldsymbol{\epsilon} \right] = \mathbf{0}$ holds, we can form an unbiased estimating function

$$\mathrm{E}\left[(\mathbf{I}_J \otimes \mathbf{a}') \, (\boldsymbol{Y} - \boldsymbol{p}) \right] = \mathbf{0} \,, \tag{1.64}$$

with sample analog

$$N^{-1} (\mathbf{I}_J \otimes \mathbf{a}') \, (\boldsymbol{Y} - \boldsymbol{p}) = \mathbf{0} \,. \tag{1.65}$$

If $N > K$ (as is often the case), the inverse problem based on this set of estimating equations is ill-posed.

One way to solve the ill-posed inverse problem and recover information about

the unknown model components is to use the CRPD criterion introduced in Section 1.3. For expositional simplicity, we focus on the special case of CRPD that results in Shannon's entropy functional. Under this information criterion, we can solve the following extremum problem

$$\max_{\boldsymbol{p}} \ -\boldsymbol{p}'\ln\left(\boldsymbol{p}\right) \ , \tag{1.66}$$

subject to the moment constraints

$$\left(\mathbf{I}_J \otimes \mathbf{a}'\right)\boldsymbol{y} = \left(\mathbf{I}_J \otimes \mathbf{a}'\right)\boldsymbol{p} \ , \tag{1.67}$$

and the additivity constraints

$$\left[\ \mathbf{I}_N \quad \mathbf{I}_N \quad \cdots \quad \mathbf{I}_N \ \right]\boldsymbol{p} = \mathbf{1} \ , \tag{1.68}$$

where the matrix on the left-hand-side is $(N \times NJ)$ and $\mathbf{1}$ is an $(N \times 1)$ unit vector.

The information theoretic solution to the inverse problem may be derived from the necessary conditions for this inverse problem. The intermediate form of the solution is

$$\widehat{p}_{ij} = \frac{\exp\left(-\mathbf{a}'_i\widehat{\boldsymbol{\lambda}}_j\right)}{\Omega_i\left(\widehat{\boldsymbol{\lambda}}\right)} = \frac{\exp\left(\mathbf{a}'_i\widehat{\boldsymbol{\beta}}_j\right)}{\Omega_i\left(\widehat{\boldsymbol{\beta}}\right)} \ , \tag{1.69}$$

where $\widehat{\boldsymbol{\lambda}}_j$ is the $(K \times 1)$ vector of optimal Lagrange multipliers for the jth moment constraint. The expression represents only an intermediate solution to the inverse problem because $\widehat{p}_{ij}$ is a function of $\widehat{\boldsymbol{\lambda}}_j$, which must be numerically determined. As indicated, the inverse problem may also be stated in terms of the response parameters $\widehat{\boldsymbol{\beta}}_j = -\widehat{\boldsymbol{\lambda}}_j$. Finally, the denominator component or partition function takes the form

$$\Omega_i\left(\widehat{\boldsymbol{\beta}}\right) = \sum_{j=1}^{J} \exp\left(\mathbf{a}'_i\widehat{\boldsymbol{\beta}}_j\right) \ . \tag{1.70}$$

Thus, by making use of the micro data in this multinomial context, we can recover estimates of the response parameters $\boldsymbol{\beta}_j$ and the corresponding marginal probabilities. Further, the solution to the inverse problem has the same mathematical form as the logistic multinomial probability model (Mittelhammer, Judge, and Miller, Chapter 20).

The intermediate solution may be substituted back into the Lagrange expression to form a concentrated objective function

$$M\left(\boldsymbol{\lambda}\right) = \boldsymbol{y}'\left(\mathbf{I}_N \otimes \mathbf{a}\right)\boldsymbol{\lambda} + \sum_{j=1}^{J}\ln\left(\Omega_i\left(\boldsymbol{\lambda}\right)\right) \ , \tag{1.71}$$

which is identical to the log-likelihood function for the multinomial logit problem

(Maddala, 1983, p. 36). Consequently, the asymptotic properties of the multinomial logit estimator also apply to the information theoretic solution in this inverse problem, and the sampling results may be used to form inferences regarding voter response to changes in the explanatory variables. The solution to the inverse problem will not coincide with the multinomial logit case if we use other members of the CRPD criterion family as the objective function. However, related large-sample properties may be derived under comparable regularity conditions.

1.4.2.1 Remarks

Precincts represent an aggregate unit, which, moreover, is aggregated at an arbitrary level. Precinct behavior is interesting in some contexts, but another challenge is reconstructing individual-level behavior. Knowing how people vote is instrumental to understanding the dynamics and impact of our political structures. Surveys provide one means of accomplishing this task. However, surveys have clear weaknesses that could be overcome with aggregate data. The discrete choice formulations developed here provide a method for utilizing survey information in conjunction with the aggregate data, and thus allow one to draw from the strengths of both levels of data. For a discussion of this type of model in an epidemiology context, see Wakefield and Salway (2001).

The discrete choice formulations provide but one way to bridge the chasm between the macro and micro estimates. They enable us to condition on a set of covariates to make this link from the macro to the micro. There have been many studies seeking to link covariates to voter choice at the micro level. We look to these studies to guide the choice of explanatory variables for the discrete choice formulations in Section 1.4. In particular, many of these studies have established a clear empirical link between voter preferences and socio-economic variables such as age, income, and education. In addition, the socio-economic variables can be used to design a survey that would elicit information on individual attitudes and how these characteristics map to attitudes. Indeed, we are more generally interested in mapping attitudes to characteristics rather than the more narrow question of how attitudes map to vote preferences. The former mapping is much more general and would allow us to engage in a wider range of prediction. Campaign strategists, after all, are most interested in forming effective targeting strategies based on individual characteristics, not individual vote preferences per se.

Although there are many ways to transform this problem into a well-posed inverse problem, our formulation here is attractive because it has many of the same nice features as the one discussed in Section 1.3. In particular, the procedure has an information theoretic–empirical likelihood base that permits semiparametric inference and allows, when available, the incorporation of non-sample information.

1.5 Implications

Ecological inference problems provide an interesting challenge for polimetricians. The secret ballot is designed to maintain an air of secrecy around individual vote preferences, and it has very successfully done so. As a result, the data generated from any election are partial and incomplete. Consequently, the corresponding estimation and inference models present themselves as underdetermined and ill-posed inverse problems. While our goal is to obtain information in terms of conditional probabilities as a basis for expressing the micro processes underlying the macro outcome data, these conditional probabilities are unobserved and unobservable. This means few, if any, bets on the values of the unknowns, will ever be collected.

Although many theories about voting behavior seem to exist, there does not appear to be one overarching micro theory that encompasses all of the empirical and theoretical research on the topic. Few have even discussed, or even seem willing to discuss, the prospects of constructing a micro foundation for aggregate outcomes. This lack of model structure creates presentational and interpretational problem and results in insufficient information on which to specify a data sampling process that might be consistent with the observed data outcomes. Hence, traditional estimation and inference procedures appear to be ill-suited to deal with ecological data. The use of creative assumptions to achieve tractability and well-posed mathematical and statistical models leads in many cases to erroneous interpretations and conclusions. No one ever said ecological inference was easy.

Building on the productive efforts of many polimetricians, in an effort to make some progress on these interesting problems and challenges, we have considered non-traditional methods of thinking about this problem. This approach recognizes that the problem of sorting out voter behavior that is modeled in terms of unknown probabilities while making use of only aggregate data constraints results in an ill-posed inverse problem. In seeking a basis for reasoning in this logically indeterminate situation, we have modeled the ecological inference problem as a pure or a noisy inverse problem. In this context, to choose a "solution" from the set of feasible solutions, the Cressie-Read statistic was used to identify a family of goodness-of-fit or psuedo distance measures. This solution provides a useful way to summarize a micro system that is consistent with the observed macro counterpart. This formulation is also attractive in that it provides a straightforward way to include prior non-sample information, is amenable to a multiplicity of precincts, can easily include spatial and intertemporal aspects, and is easy to implement. This approach also allows one to alleviate problems such as measurement error by incorporating an instrumental variables framework that may be employed along with the moment conditions to provide a basis for recovery estimates of response parameters and the corresponding marginal probabilities related to voter preferences. Finally, in order to bridge these estimates to the micro-level, we view

the ecological inference problem as a discrete choice problem. This permits the recovery of response parameters related to voter characteristics, and again recovery of corresponding marginal (conditional) probabilities. It is worth noting that the application of maximum entropy methods has been explored in the ecological inference context (Johnston and Pattie 2002). However, extensions of the method which we explore (i.e. cross-entropy) and the introduction of information theoretic techniques is novel to the study of ecological inference.

Under the instrumental variables and discrete choice formations, estimation and inference proceeds in the context of sampling theory and provides a sampling basis for evaluating performance. To a large extent, the information processing and recovery rules described are non-traditional in nature and does not assume information about the underlying sampling distributions, which is unknown. These nonparametric/semiparametric formations permit one to stay within the realm of sampling theory but allow one to avoid the rigidity of likelihood functions and proceed based only on a finite set of moment conditions.

In looking ahead in terms of ways to think about ecological inference problems, semiparametric and nonparametric formulations of the random coefficient models seem to be promising avenues. In this framework, one may replace unknown functions with reasonable nonparametric estimators rather than the maximum likelihood estimator that constrains the parametric setting. One possibility in this connection is sieve empirical likelihood estimation and testing procedures. Alternatively, the Bayesian method of moments offers a basis for recovering conditional probabilities without the usual Bayesian likelihood and prior distributions.

The writing of this paper, which led to a trek into the world of ecological inference, has been a very rewarding experience. It has reminded us that aggregate analyses that lead to invalid micro inferences also have implications and consequences other than those of the statistical ilk. It is also refreshing for economists to be reminded that the problem of recovering micro level effects from an aggregate counterpart is not unique to economic data.

1.6 References

L. Anselin and W.K.T. Cho, "Spatial Effects and Ecological Inference," Political Analysis, Vol. 10, No. 3, pp. 276–297, 2002.

K. Baggerly, "Empirical Likelihood as a Goodness of Fit Measure," Biometrika, Vol. 85, No. 3, pp. 535–47, 1998.

A. Bera and Y. Bilias, "The MM, ME, ML, EL, EF, and GMM Approaches to Estimation: a synthesis," Journal of Econometrics, Vol. 107, pp. 51–86, 2002.

E. Calvo and M. Escolar, "The Local Voter: Exploring a Geographically Weighted Approach to Cross-level Inference," Paper presented at the Harvard Ecological Inference Conference, 2002.

S. Corcoran, "Empirical Exponential Family Likelihood using Several Moment Conditions," Statistica Sinica, Vol. 10, pp. 545–57, 2000.

N. Cressie and T. Read, "Multinomial Goodness of Fit Tests," Journal of the Royal Statistical Society, Series B, Vol. 46, pp. 440–64, 1984.

I. Csiszár, "I–divergence Geometry of Probability Distributions and Minimization Problems," Annals of Probability, Vol. 3, pp. 146–58, 1975.

I. Csiszár, "Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems," Annals of Statistics, Vol. 19, pp. 2032–66, 1991.

W. E. Deming and F. Stephan, "On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known," Annals of Mathematical Statistics, Vol. 11, pp. 423–44, 1940.

T. DiCicco and J. Romano, "Nonparametric Confidence Limits by Resampling Methods and Least Favorable Families," International Statistics Review, Vol. 58, pp. 59–76, 1990.

O. Duncan and B. Davis, "An Alternative to Ecological Correlation," American Sociological Review, Vol. 18, pp. 665–6, 1953.

D. Freedman, M. Ostland, M. Roberts, and S. Klein, "Response to King," Journal of the American Statistical Association, Vol. 94, pp. 355–7, 1999.

D. Freedman, S. Klein, J. Sacks, C. Smyth, and C. Everett, "Ecological Regression and Voting Rights (with discussion)," Evaluation Review, Vol. 15, pp. 659–817, 1991.

D. Freedman, S. Klein, M. Ostland, and M. Roberts, "Review of 'A Solution to the Ecological Inference Problem'," Journal of the American Statistical Association, Vol. 93, pp. 1518–22, 1998.

D. Gokhale and S. Kullback, The Information in Contingency Tables, Marcel Dekker, 1978.

A. Golan, G. Judge, and D. Miller, Maximum Entropy Econometrics: Robust Estimation with Limited Information, John Wiley and Sons, 1996.

A. Golan, G. Judge, and S. Robinson, "Recovering Information from Incomplete or Partial Multisectoral Economic Data," Review of Economics and Statistics, Vol. 76, pp. 541–9, 1994.

I. J. Good, "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables," Annals of Mathematical Statistics, Vol. 34, pp. 911–34, 1963.

L. Goodman, "Ecological Regression and the Behavior of Individuals," American Sociological Review, Vol. 18, pp. 663–4, 1953.

L. Goodman, "Some Alternatives to Ecological Correlation," American Journal of Sociology, Vol. 64, pp. 610–25, 1959.

C.A. Gotway and L.J. Young, "A Spatial View of the Ecological Inference Problem," Paper presented at the Harvard Ecological Inference Conference, 2002.

B. Grofman, L. Handley, and R. Niemi, Minority Representation and the Quest for Voting Equality, Cambridge University Press, 1992.

S. Haberman, "Adjustment by Minimum Discriminant Information," Annals of Statistics, Vol. 12, pp. 971–88, 1984.

S. Haneuse and J. Wakefield, "Ecological Inference Allowing for Spatial Dependence," Paper presented at the Harvard Ecological Inference Conference, 2002.

C. Ireland and S. Kullback, "Contingency Tables with Given Marginals," Biometrika, Vol. 55, pp. 179–88, 1968.

C. Ireland and S. Kullback, "Minimum Discriminant Information Estimation," Biometrics, Vol. 24, pp. 707–13, 1968.

E. T. Jaynes, "Information Theory and Statistical Mechanics I," Physics Review, Vol. 106, pp. 620–30, 1957.

E. T. Jaynes, "Information Theory and Statistical Mechanics II," Physics Review, Vol. 108, pp. 171–90, 1957.

R. Johnston and A. Hay, "Voter Transition Probability Estimates: An Entropy–Maximizing Approach," European Journal of Political Research, Vol. 11, pp. 93–98, 1983.

R. Johnston and C. Pattie, "Ecological Inference and Entropy–Maximizing: An Alternative Estimation Procedure for Split–Ticket Voting," Political Analysis, Vol. 8, pp. 333-45, 2000.

G. King, A Solution to the Ecological Inference Problem, Princeton University Press, 1997.

G. King, "The Future of Ecological Inference Research: A Reply to Freedman, et al.," Journal of the American Statistical Association, Vol. 94, pp. 352–55, 1999.

J. Kousser, 1973. "Ecological Regression and the Analysis of Past Politics," The Journal of Interdisciplinary History, 4 (Autumn),2: 237–262, 1973.

S. Kullback, Information Theory and Statistics, John Wiley and Sons, 1959.

S. Kullback, "Marginal Homogeneity of Multidimensional Contingency Tables," Annals of Mathematical Statistics, Vol. 42, pp. 594–606, 1971.

G. S. Maddala, Limited–Dependent and Qualitative Variables in Economics, Cambridge University Press, 1983.

C. Manski and D. McFadden, "Alternative Estimates and Sample Designs for Discrete Choice Analysis," In C. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, The MIT Press, 1982.

K. McCue, "The Statistical Foundations of the EI Method," The American Statistician, Vol. 55, pp. 106–10, 2001.

D. McFadden, "The Measurement of Urban Travel Demand," Journal of Public Economics, Vol. 3, pp. 303–28, 1974.

R. Mittelhammer, G. Judge, and D. Miller, Econometric Foundations, Cambridge University Press, 2000.

J. Neyman, "Contributions to the Theory of the χ^2 Test," Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability, pp. 239–73, 1949.

A. Owen, "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," Biometrika, Vol. 75, pp. 237–49, 1988.

A. Owen, "Empirical Likelihood Ratio Confidence Regions," Annals of Statistics, Vol. 18, pp. 90–120, 1990.

K. Pearson, "On a criterion that a given system of deviations from the probable in the case

of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," Philosophical Magazine, 5th Series, Vol. 50, pp. 157–75, 1900.

T. Read and N. Cressie, Goodness-of-Fit Statistics for Discrete Multivariate Data, Springer-Verlag, 1988.

C. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal, Vol. 27, pp. 379–423, 1948.

J. Wakefield and R. Salway, "A Statistical Framework for Ecological and Aggregate Studies." Journal of the Royal Statistical Society, Series A, Vol. 164, pp. 119–137, 2001.

Appendix A: Illustrative Examples

Information-Theoretic Formulation

To illustrate the properties of the information theoretic approach, we consider a special case of the VPI problem based on an election with $k = 4$ candidates in some precinct. Suppose that N votes were cast for the candidates, and that we record the individual votes as $x_i = j$ for $i = 1, \ldots, N$ and $j = 1, \ldots, 4$. Because of the secret ballot, individual records are unknown and thus we only have the average vote outcome from the election, $\bar{x}$. Further, suppose we believe that the candidates are equally likely to win the election *ex ante*. The objective of our VPI problem is to estimate the proportion β_k of votes that each candidate received based on this very limited information. Within the context of Section 1.3, we solve the problem by maximizing the CRPD objective function with uniform reference weights and $\lambda \to 0$

$$ -\sum_{k=1}^{4} \beta_k \, \ln\left(\beta_k\right) \; , \tag{A.1} $$

subject to

$$ \sum_{k=1}^{4} \beta_k x_k = \bar{x} \; , \tag{A.2} $$

$$ \sum_{k=1}^{4} \beta_k = 1 \; , \tag{A.3} $$

by choice of $\beta_k \geq 0$. The intermediate solution to the VPI problem is

$$ \widehat{\beta}_k = \frac{\exp\left(-\widehat{\alpha} x_k\right)}{\sum_{k=1}^{4} \exp\left(-\widehat{\alpha} x_k\right)} \; , \tag{A.4} $$

where $\widehat{\alpha}$ is the optimal Lagrange multiplier for the constraint (A.2).

Although the problem is stated as a constrained maximization, the computational burden may be reduced by concentrating the estimation problem. Following the discussion of (1.71) for the discrete choice problem in Section 1.4.2, we can substitue the intermediate solution (A.4) back into the Lagrange equation for the problem defined by (A.1) to (A.3). The resulting concentrated objective function

$$ M\left(\alpha\right) = \alpha \bar{x} + \ln\left[\sum_{k=1}^{4} \exp\left(-\alpha x_k\right)\right] \; , \tag{A.5} $$

is strictly convex in α, and the optimal value of the Lagrange multiplier may be computed by minimizing $M\left(\alpha\right)$. We can then evaluate (A.4) at $\widehat{\alpha}$ to determine the estimated vote shares. Thus, the estimates for the VPI problem may be computed with any software package that solves unconstrained optimization problems (e.g.,

Table A.1 *Solutions to the Simple VPI Problem*

	Candidate			
$\bar{x}$	1	2	3	4
1.5	0.648	0.235	0.086	0.031
2.0	0.421	0.277	0.182	0.120
2.5	0.250	0.250	0.250	0.250
3.0	0.120	0.182	0.277	0.421
3.5	0.031	0.086	0.235	0.648

maximum likelihood or nonlinear least square estimation). In general, we can form concentrated objective functions for any of the minimum CRPD problems stated in this chapter, and we provide additional examples in the following subsection.

The simple VPI problem is very similar to Jaynes' famous dice problem in which we must assign probabilities to the six faces of a die based on the observed average outcome of N rolls. In our case, we have four unknown probabilities β_k and only two pieces of available information. To demonstrate the solution to our simple VPI problem, we report the conditional probabilities for five different values of $\bar{x}$ in Table A.1. Note that all of the minimum CRPD solutions to this problem based on uniform reference weights are discrete uniform when $\bar{x} = 2.5$. Otherwise, the estimated conditional probabilities are monotonically increasing if $\bar{x} > 2.5$ and monotonically decreasing if $\bar{x} < 2.5$.

Variants of the King Ohio Voter Problem

To further demonstrate the minimum CRPD procedure, we consider the simple problem presented by King (1997) in his Table 1.2. For a particular Ohio precinct, King reports the number of votes for the two major parties plus the number of non-voters ($c = 3$) in the 1990 Ohio State House election. King also reports the number of registered black and white voters in the Ohio precinct ($g = 2$). The data provided by King are the row and column sums in Table A.2. For example, there are 221 black registered voters in the precinct, and 92 votes were cast for the Republican candidate. The associated VPI problem is to estimate the number of votes cast for each party (including no-votes) conditional on the race of the voter. In effect, we have $g(c - 1) = 4$ unknown probabilities and $(g - 1) = 2$ pieces of information (after normalization), and King's Ohio voter problem is clearly underdetermined.

The Ohio voter problem is solved using the minimum CRPD estimator with uniform reference weights and $\lambda \to 0$. The objective function is (1.23), and the

intermediate solution for the constrained optimal β_{ijk} is a special case of (1.22)

$$\widehat{\beta}_{jk} = \frac{\exp\left(-\widehat{\alpha}_k n_{j\cdot}\right)}{\sum_{k=1}^{3} \exp\left(-\widehat{\alpha}_k n_{j\cdot}\right)} \ . \tag{A.6}$$

To compute the optimal Lagrange multipliers $\widehat{\alpha}_k$, we minimize the concentrated objective function

$$M\left(\boldsymbol{\alpha}\right) = \sum_{k=1}^{3} n_{\cdot k} \alpha_k + \sum_{j=1}^{2} \ln\left[\sum_{k=1}^{3} \exp\left(-\alpha_k n_{j\cdot}\right)\right] \ . \tag{A.7}$$

The predicted vote counts appear in the individual cells in the table, and the estimated conditional probabilities are reported in parentheses. Without access to the individual ballots, we cannot know the true values of the β_{ijk} elements in this example. However, we do know that the solution is consistent with a reasonable set of regularity conditions and with what is known about the set of feasible conditional probabilities, β_{ijk}. Further, the estimated voter counts have maximum multiplicity under the Shannon entropy criterion. That is, the conditional distribution that maximizes (1.23) is coincident with the set of cell-specific vote outcomes that may be realized in the largest number of ways given the row and column sum constraints (see Section 2.2 in Golan, Judge, and Miller for more details).

To demonstrate the impact of the bounds on β_{ijk}, we solve the bounded VPI problem and present the results in Table A.2. For this version of the VPI problem stated in Section 1.3.3, the weights on the bounds are special cases of (1.31)

$$\widehat{\varphi}_{jkh} = \frac{\exp\left(-\widehat{\alpha}_k n_{j\cdot} Z_{jkh} - \widehat{\gamma}_j Z_{jkh}\right)}{\sum_{h=1}^{2} \exp\left(-\widehat{\alpha}_k n_{j\cdot} Z_{jkh} - \widehat{\gamma}_j Z_{jkh}\right)} \ , \tag{A.8}$$

The concentrated objective function for this problem

$$\begin{aligned} M\left(\boldsymbol{\alpha}, \boldsymbol{\gamma}\right) &= \sum_{k=1}^{3} \alpha_k n_{\cdot k} + \sum_{j=1}^{g} \gamma_j \\ &+ \sum_{j=1}^{2}\sum_{k=1}^{3} \ln\left[\sum_{h=1}^{2} \exp\left(-\alpha_k n_{j\cdot} Z_{jkh} - \gamma_j Z_{jkh}\right)\right] \ , \end{aligned} \tag{A.9}$$

is derived by substituting the intermediate solution back into the Lagrange equation, and the optimal Lagrange multipliers $\widehat{\alpha}_k$ and $\widehat{\gamma}_j$ are computed by unconstrained minimization of $M\left(\boldsymbol{\alpha}, \boldsymbol{\gamma}\right)$. The solution values of β_{ijk} are computed from the optimal weights $\widehat{\varphi}_{jkh}$ as in (1.29), and the estimates appear in parentheses below the estimated vote counts. The associated Duncan-Davis bounds appear below estimate in brackets. In five of six cells, the upper or lower bounds narrow the feasible set to a proper subset of [0,1]. Also, the estimated conditional probabilities are near (but not exactly at) the center of the bounded intervals. Relative to the unbounded results, note that the bounding information has effectively shifted votes from the no-vote category to the Democrat (Republican) columns for black

(white) voters. Of course, we do not know that this solution is better than the unbounded pure solution because the cell values are unobservable.

To illustrate the case with bounds and $\lambda \to 0$ from Section 1.3.4.2, we solve the Ohio voting example in terms of the extended VNI problem with bounds on β_{ijk} and ε_{ik}. We use the bounds on β_{ijk} stated in Table A.2, and the upper and lower bounds appear below the estimates in the table. The error bounds are selected to be symmetric about zero with $\delta_{ik} = 1$. The upper and lower error bounds are stated below the column counts at the bottom of Table A.2. The intermediate solution for φ_{jkh} takes the same form as in the bounded VPI problem, but the optimal values of α and γ for this problem will be different due to the presence of the noise term. The concentrated objective function is

$$M^*(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = M(\boldsymbol{\alpha}, \boldsymbol{\gamma}) + \sum_{k=1}^{3} \ln\left[\exp\left(-\alpha_k V_{k1}\right) + exp\left(-\alpha_k V_{k2}\right)\right] , \quad (A.10)$$

and the term added to $M(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ represents the presence of the noise terms. Relative to the two preceding demonstrations, note that the estimates for the bounded VNI problem represents an intermediate case—some of the mass shifted to form the bounded VPI estimates has reverted to the unbounded VPI case. In effect, the noise components weaken the constraints for the VNI problem, and the solution can move closer to the unbounded outcome. Although the column sums are not strictly required to match the observed values, note that this property is satisfied by this solution. Further, the use of wider error bounds reduces the tendency for the column sums to be satisfied by the estimated conditional probabilities.

This is a fairly simplistic example that could be extended easily in several ways. For instance, under the usual scenarios, candidates are far from equally likely to win the election *ex ante*. This assumption can be weakened so that we can incorporate our fairly accurate ability to predict election outcomes much before they occur. In an actual application of this problem, we will be able to capitalize on the information provided by a larger number of precincts. This example supplies estimates for just one precinct. Presumably, the numerous precincts that would comprise a data set would supply additional information. We could perhaps take advantage of information underlying some manifested spatial autocorrelation among the precincts (see work in this area, Anselin and Cho 2002, Calvo and Escolar 2002, Gotway and Young 2002, and Haneuse and Wakefield 2002).

Given the large degree of uncertainty that surrounds these estimates, it is difficult to choose between these three demonstrations. One might be inclined to tend toward the VNI formulation with bounds simply because the bounds are deterministic information that one would like to incorporate and the errors certainly seem to be important and plausible features as well. Interestingly, however, none of these demonstrations produces substantively different results. And even without explicitly incorporating the bounds, the estimates for the VPI problem are

within the bounds. This lack of variation in the estimates is especially true for the Republican candidate where the range of vote counts is minuscule (45–47) as is the range of vote proportions. The range for the Democratic candidate is larger, but still not big enough to be substantively interesting. The bounds in this case are relatively narrow, so it is especially surprising that they would not have a greater effect.

Lastly, we note that the estimates in these cases, especially when the bounds are incorporated, appear to tend toward the "center" of the possible range of values, and that the estimates for the white group and the black group tend to be more similar than not. This is not particularly surprising, as one might initially guess that the estimated voter counts would have maximum multiplicity toward the center rather than toward either end of the range of possibilities. In this sense, one might believe that this estimator is more inclined to imply that different groups of voters tend to act more similarly rather than less similarly. The implications for using this estimator in a Voting Rights case, then, can be quite consequential, since the charge in those cases is to determine whether there is polarized voting among the groups in the electorate and the inclination of the estimator is to provide group estimates that are more similar than not.

Moment-Based Formulation

We now demonstrate the sampling properties of the estimator for the moment-based model formulation in Section 1.4.1. In particular, we consider a cross-sectional version of the model based on $m = 20$ and $m = 50$ precincts with $g = 3$ voter types and $c = 4$ candidates. Three instrumental variables $\mathbf{A}_i$ are used for each precinct, and these include a constant (i.e., vector of ones) and two non-constant instruments generated as pseudo-random standard normal variables (fixed in repeated samples). We also simplify the formulation stated in Section 1.4.1 by using uniform reference weights for the conditional probabilities and by ignoring the bounds on the conditional probabilities. Following the notation in (1.57), we denote the associated Lagrange multipliers for this version of the model as α_{kh} where $k = 1, \ldots, 4$ and $h = 1, \ldots, 3$. The row-sum values are the same for each precinct, $n_{i1\cdot} = 0.3$, $n_{i2\cdot} = 0.25$, and $n_{i3\cdot} = 0.45$. The mean values of the column-sum values $n_{\cdot i k}$ are derived from (1.1) based on a set of "true" conditional probabilities that are functions of the observed instrumental variables $\mathbf{A}_i$ and the true values of the Lagrange multipliers α_{kh} (see Table A.3 for the true values). To represent sampling variability in the candidate shares as in (1.5), we add Gaussian noise components with mean zero and variance 0.0001 to the mean values of $n_{\cdot i k}$ for $k = 1, 2, 3$. The noisy value of $n_{\cdot i 4}$ is recovered by normalization, $n_{\cdot i 4} = 1 - n_{\cdot i 1} - n_{\cdot i 2} - n_{\cdot i 3}$.

The sampling process is replicated for $m = 20$ and $m = 50$ precincts over 500 Monte Carlo trials. The estimated Lagrange multipliers are saved from each trail, and the sample mean and standard deviation of the estimates for each α_{kh} are

Table A.2 *Estimates for the Ohio Voter Problem*

VPI Problem

| | Candidate | | | |
Group	Democrat	Republican	No-Vote	Count
black	56.8	46.0	118.2	221
	(0.2570)	(0.2080)	(0.5350)	
white	73.2	46.0	364.8	484
	(0.1512)	(0.0951)	(0.7536)	
	130	92	483	705

VPI Problem with Bounded Probabilities

| | Candidate | | | |
Group	Democrat	Republican	No-Vote	Count
	67.0	47.0	107.0	
black	(0.3030)	(0.2130)	(0.4840)	221
	[0, 0.588]	[0, 0.416]	[0,1]	
	63.0	45.0	376.0	
white	(0.1300)	(0.0930)	(0.7770)	484
	[0, 0.269]	[0, 0.190]	[0.541, 0.998]	
	130	92	483	705

VNI Problem with Bounded Probabilities and Errors

| | Candidate | | | |
Group	Democrat	Republican	No-Vote	Count
	64.9	45.9	110.2	
black	(0.2940)	(0.2080)	(0.4990)	221
	[0, 0.588]	[0, 0.416]	[0,1]	
	65.1	46.1	372.8	
white	(0.1350)	(0.0950)	(0.7700)	484
	[0, 0.269]	[0, 0.190]	[0.541, 0.998]	
	130	92	483	705
	[-0.031, 0.031]	[-0.033, 0.033]	[-0.026, 0.026]	

Table A.3 *Simulation Results for the Moment-Based Formulation*

Parameter	True value	$m = 20$		$m = 50$	
		Average	Std. dev.	Average	Std. dev.
α_{11}	0.1	0.1288	0.94	0.1378	1.13
α_{12}	0.2	0.1559	1.49	0.1890	1.64
α_{13}	0.3	0.2795	1.87	0.3002	0.25
α_{21}	0.4	0.4296	0.95	0.4385	1.12
α_{22}	0.5	0.4580	1.49	0.4886	1.63
α_{23}	0.6	0.5998	1.87	0.5997	0.24
α_{31}	0.7	0.7276	0.94	0.7381	1.13
α_{32}	0.8	0.7570	1.49	0.7890	1.64
α_{33}	0.9	0.8826	1.87	0.9001	0.26
α_{41}	1.0	1.0303	0.95	1.0394	1.13
α_{42}	1.1	1.0578	1.50	1.0856	1.62
α_{43}	1.2	1.1795	1.86	1.1988	0.23

reported with the true parameter values in Table A.3. Given that this is a cross-sectional sample $(T = 1)$, the regularity conditions stated in Section 1.4.1 do not directly apply to this case. However, we find that the sample means of the $\hat{\alpha}_{kh}$'s are relatively close to the true parameter values and the simulated standard errors are stable. The sampling results are especially encouraging because we are not relying on time series observations $(T = 1)$ and the largest value of m is small relative to typical number of precincts. Further, the relative efficiency of the moment-based estimator may be improved by accounting for spatial correlation among the precincts.