Journal of the
Royal Statistical Society

SERIES A
Statistics
in Society

J. R. Statist. Soc. A (2009)
172, *Part* 1, *pp.* 67–81

$R \times C$ ecological inference: bounds, correlations, flexibility and transparency of assumptions

D. James Greiner

Harvard Law School, Cambridge, USA

and Kevin M. Quinn

Harvard University, Cambridge, USA

[Received July 2006. Final revision February 2008]

Summary. Despite its potential pitfalls, ecological inference is an unavoidable part of some quantitative settings, including US voting rights litigation. In such applications, the analyst will typically encounter two-way tables with more than two rows and columns. Although several ecological inference methods are currently available for 2×2 tables, there are fewer options for analysing general $R \times C$ tables, and virtually none that model counts as opposed to fractions. We propose a count $R \times C$ method that respects the bounds deterministically, that allows for complex relationships between internal cell quantities, that is easily extensible and that results from transparent assumptions. We study the method via simulation, and then apply it to an example that is drawn from the state of Texas relevant to recent redistricting litigation there.

Keywords: Bayesian inference; Ecological inference; Voting rights litigation

1. Introduction and basic notation

Ecological inference, or the effort to draw conclusions about statistical relationships at one level from data that are aggregated to a higher level, has a long history in statistics. Particularly when conceptualized in terms of an effort to predict the internal values of a set of contingency tables when only the row and column totals are observed, ecological inference has well-known pitfalls, primarily that substantively different internal cell counts can give rise to the same marginal totals (Robinson, 1950). This potential has led some to the respectable view that this form of inference should never be attempted (e.g. Freedman *et al.* (1991)), but in certain applications, including US redistricting litigation, there is little alternative (see Greiner (2007)).

To date, the overwhelming majority of research on ecological inference has focused on a set of contingency tables with two rows and two columns, but in many settings (including redistricting) the relevant tables are usually larger. For this reason, we build on work by Brown and Payne (1986) and Wakefield (2004) to propose a model and a corresponding fitting algorithm for $R \times C$ tables. Our method may be seen as a generalization of a model for 2×2 tables that was articulated by Wakefield (2004), although not the generalization that Wakefield himself proposed. The advantages of our method include the following:

Address for correspondence: D. James Greiner, Harvard Law School, 504 Griswold Hall, Cambridge, MA 02138, USA.
E-mail: jgreiner@law.harvard.edu

Journal compilation © 2009 Royal Statistical Society 0964–1998/09/172067

Table 1. 3 × 3 table of voting by race

	Democrat	*Republican*	*Abstain*	
Black	N_{bD_i}	N_{bR_i}	N_{bA_i}	$\mathrm{N}_{\mathrm{b}_\mathrm{i}}$
White	N_{wD_i}	N_{wR_i}	N_{wA_i}	$\mathrm{N}_{\mathrm{w}_\mathrm{i}}$
Hispanic	N_{hD_i}	N_{hR_i}	N_{hA_i}	$\mathrm{N}_{\mathrm{h}_\mathrm{i}}$
	$\mathrm{N}_{\mathrm{D}_\mathrm{i}}$	$\mathrm{N}_{\mathrm{R}_\mathrm{i}}$	$\mathrm{N}_{\mathrm{A}_\mathrm{i}}$	N_i

(a) deterministic respect for the bounds (see Duncan and Davis (1953)) within a parametric structure (see King (1997)) that permits coherent expression and exploration of prior assumptions;
(b) correspondence to a well-specified model at the level of the units subject to aggregation (voters, in redistricting), which facilitates incorporation of survey information, assessment of assumptions and explanation of the method to lay audiences;
(c) a focus on counts instead of fractions;
(d) allowance for complex relationships among counts inside the contingency tables;
(e) flexibility to explore a variety of extensions.

We note, however, that the basic form of our model, as is true of almost all ecological inference methods, is not immune to aggregation bias.

We organize this paper as follows. We briefly discuss existing $R \times C$ ecological inference techniques before listing our goals in formulating our proposal. We then outline our method (including its motivation), identify quantities of interest, examine priors, present simulation results and apply the method to a data set. We continue with a discussion of extensions of the method before concluding.

To motivate the quantitative problem, we suppose that we seek inferences about the voting behaviour of different racial (racial or ethnic) groups, and we observe the data aggregated to the level of the precinct. We use the symbol $N_{\mathrm{row\ COLUMN}_i}$ to refer to the (unobserved) count in a particular cell of the ith precinct, where rows represent races and columns represent political parties (and non-votes); i runs from 1 to I, the number of precincts in the jurisdiction, and there are R rows and C columns in each precinct's table. We italicize unobserved counts but leave observed quantities in ordinary type. For illustrative purposes, we focus on the case of 3×3 precinct tables; extension of the method that we propose to tables of different size and shape is obvious. Table 1 clarifies our representations.

Some models (although not ours) work with the fractional quantities that are generated by dividing each cell by its corresponding row total. We label unobserved internal cell fractions $\beta_{\mathrm{row\ COLUMN}_i}$; so, for example, $\beta_{bD_i} = N_{bD_i}/\mathrm{N}_{\mathrm{b}_i}$. $\mathrm{X}_{\mathrm{row}_i}$ refers to the racial population share of the ith precinct, so $\mathrm{X}_{h_i} = \mathrm{N}_{\mathrm{h}_\mathrm{i}}/\mathrm{N}_\mathrm{i}$.

2. Existing $R \times C$ techniques, and goals

One classification of existing $R \times C$ ecological inference techniques separates methods that eliminate or collapse rows and/or columns from methods that estimate all internal cells simultaneously. Elimination of columns (e.g. the 'single regression' of Grofman *et al.* (1985)) or rows (e.g. Benoit *et al.* (2004)) proceeds via an undesirable assumption that cells corresponding to a particular category exhibit no systematic patterns. Collapsing methods (e.g. King (1997)) typically involve successive application of 2×2 techniques to sets of tables that are formed by combining row and/or column categories to form supertotals (see Ferree (2004) for a discussion of the drawbacks of at least one collapsing technique). We prefer methods that estimate all cells of precinct tables simultaneously.

A different classification of existing methods turns on whether they predict fractions (βs, in our notation) or counts (Ns). The distinction is subtle but important. Models in the former category include the linear model (whether constrained, *per* Gelman *et al.* (2001), or unconstrained; see Achen and Shively (1995)), King's (1997) ecological inference, the hierarchical method of Rosen *et al.* (2001) and the information theoretic proposal in Judge *et al.* (2004). Our uncertainty with this approach begins with the interpretation and use of the βs, particularly in light of the fact that, in many applications, quantities of interest are functions of the internal cell counts (the Ns) in the precinct tables. For example, Judge *et al.* (2004) (who also used redistricting examples) characterized the βs as row probabilities. If the βs are row probabilities, however, it would seem that inference about the internal cell counts would require specification of a separate distribution of the counts given these probabilities (such as a multinomial), but Judge *et al.* (2004) instead suggested that predictions for the internal counts should be calculated deterministically as the product of the (observed) row totals and the βs.

The choice to work with counts *versus* fractions also constitutes a fundamental decision about how to treat precincts of varying size. For example, most fraction models would treat precinct Tables 2 and 3 (which we make 2×2 to illustrate our point) as having an identical amount of information about the turnout of blacks. Although the concept is difficult to quantify rigorously, models that work with counts would treat Tables 2 and 3 differently in that Table 3 might have greater information on black turnout behaviour than Table 2. Such treatment might be preferable depending in part on whether the analyst wishes to build a model from the individual voter upwards.

The discussion above rests on certain *desiderata* for an $R \times C$ ecological inference method; we now list these explicitly. First, we seek an $R \times C$ method that respects the bounds on the interior cell counts deterministically. Second, we seek a method that works with counts (as opposed to

Table 2. 2 × 2 table: 100 eligible to vote

	Vote	*No vote*	*Total*
Black			96
White			4
	50	50	100

Table 3. 2 × 2 table: 10000 eligible to vote

	Vote	*No vote*	*Total*
Black			9600
White			400
	5000	5000	10000

fractions) inside the precinct tables, a choice which provides a closer fit to the data-generating process (see, for example, Prentice and Sheppard (1995) in epidemiology) by means of an individual level model of voting behaviour. Such a model facilitates communication and assessment of modelling assumptions to other scholars and to lay audiences while easing the incorporation of individual level survey information. Third, we seek a technique that is sufficiently flexible to explore within-row and between-row relationships. In the redistricting context, for example, relationships between voter choices in electoral contests involving more than one candidate of each race may shed light on whether voting in the jurisdiction is cued on the basis of race. The model that we propose is sufficiently flexible on this score, and we use simulation to study how the aggregation process causes differing amounts of loss of information for between- and within-row relationships. Fourth, an $R \times C$ technique should allow for a variety of extensions.

3. Our proposal

3.1. The individual level model and basic structure

We begin with our individual level model. We take precinct boundaries and the racial composition of each precinct as fixed, thus assuming that the way in which precinct boundaries are drawn is unrelated to the data-generating process. We then suppose that each potential voter has a probability of supporting the Democrat, of supporting the Republican or of abstaining from voting. In the basic form of the model, the potential voter's probability vector depends on exactly two things: his or her race and the precinct in which he or she lives. Individual voting decisions are mutually independent. This motivation results in a product multinomial complete-data likelihood at the bottom level of the hierarchy (see Brown and Payne (1986)).

At the second level of the hierarchy, we apply a logistic transformation to the multinomial probability vectors within each precinct, choosing the 'abstain' column as our reference category (see Aitchison (2003)). The transformation results, for each precinct, in a set of R vectors, each having support in the $(C-1)$-dimensional Euclidean space of real numbers. At this point, when discussing $R \times C$ precinct level tables, Wakefield (2004) proposed a 'simple model' that assumes that the R vectors all come from the same $(C-1)$-dimensional normal distribution. Our difficulty with this approach is that, in the voting rights context, it would mean that black, white and Hispanic voting patterns are the same. When considering 2×2 precinct level tables, Wakefield articulated a bivariate normal prior with an off-diagonal term to model dependence (see Wakefield (2004), section 5.4). This is a superior approach, and we generalize it here.

Thus, we stack the R-vectors (each of dimension $C-1$) into an overall precinct vector of transformed probabilities and assume that each stacked vector is an independent manifestation of an $R(C-1)$-dimensional normal distribution. The result is what we believe to be a previously unproposed extension of the additive logistic normal model from Aitchison (2003). For computational convenience, we put semiconjugate priors on the mean and covariance matrix of the normal distribution. In symbols, for the case of precinct tables of Table 1 form, our proposal has the following structure: level 1,

$$(N_{bD_i}, N_{bR_i}, N_{bA_i})|\mathrm{N_{b_i}}, \boldsymbol{\theta}_{b_i} \sim \mathrm{Multi}\{\mathrm{N_{b_i}}, \boldsymbol{\theta}_{b_i} = (\theta_{bD_i}, \theta_{bR_i}, \theta_{bA_i})^{\mathrm{T}}\},$$

mutually $\perp\!\!\!\perp$ of

$$(N_{wD_i}, N_{wR_i}, N_{wA_i})|\mathrm{N_{w_i}}, \boldsymbol{\theta}_{w_i} \sim \mathrm{Multi}\{\mathrm{N_{w_i}}, \boldsymbol{\theta}_{w_i} = (\theta_{wD_i}, \theta_{wR_i}, \theta_{wA_i})^{\mathrm{T}}\},$$

mutually $\perp\!\!\!\perp$ of

$$(N_{hD_i}, N_{hR_i}, N_{hA_i})|\mathrm{N_{h_i}}, \boldsymbol{\theta}_{h_i} \sim \mathrm{Multi}\{\mathrm{N_{h_i}}, \boldsymbol{\theta}_{h_i} = (\theta_{hD_i}, \theta_{hR_i}, \theta_{hA_i})^{\mathrm{T}}\},$$

mutually $\perp\!\!\!\perp$ of . . . ; level 2, setting

$$\boldsymbol{\omega}_{b_i}^{\mathrm{T}} = \left(\log\left(\frac{\theta_{bD_i}}{\theta_{bA_i}} \right), \log\left(\frac{\theta_{bR_i}}{\theta_{bA_i}} \right) \right),$$

and similarly for $\boldsymbol{\omega}_{w_i}$ and $\boldsymbol{\omega}_{h_i}$,

$$\boldsymbol{\omega}_i = (\boldsymbol{\omega}_{b_i}^{\mathrm{T}} \ \boldsymbol{\omega}_{w_i}^{\mathrm{T}} \ \boldsymbol{\omega}_{h_i}^{\mathrm{T}})^{\mathrm{T}} | \boldsymbol{\mu}, \boldsymbol{\Sigma} \overset{\mathrm{IID}}{\sim} N_6 \left\{ \boldsymbol{\mu} = (\boldsymbol{\mu}_b^{\mathrm{T}} \ \boldsymbol{\mu}_w^{\mathrm{T}} \ \boldsymbol{\mu}_h^{\mathrm{T}})^{\mathrm{T}}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bw} & \boldsymbol{\Sigma}_{bh} \\ \boldsymbol{\Sigma}_{bw} & \boldsymbol{\Sigma}_w & \boldsymbol{\Sigma}_{wh} \\ \boldsymbol{\Sigma}_{bh} & \boldsymbol{\Sigma}_{wh} & \boldsymbol{\Sigma}_h \end{pmatrix} \right\};$$

level 3,

$$\boldsymbol{\mu} | \boldsymbol{\mu}_0, \mathbf{K}_0 \sim N(\boldsymbol{\mu}_0, \mathbf{K}_0),$$
$$\boldsymbol{\Sigma} | \nu_0, \boldsymbol{\Psi}_0 \sim \mathrm{InvWish}_{\nu_0}(\boldsymbol{\Psi}_0).$$

In block matrix form, subparts of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are interpretable as governing relationships within and between precinct table rows. For example, if $\boldsymbol{\Sigma}_{bw} = \boldsymbol{\Sigma}_{bh} = \boldsymbol{\Sigma}_{wh} = \mathbf{0}$, well-known properties of the multivariate normal distribution imply that the rows of each precinct table are conditionally independent. This special case of our method bears a structural resemblance to the proposal in Rosen *et al.* (2001), although within-row relationships (which are governed by $\boldsymbol{\Sigma}_b$, $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}_h$) are less constrained in the former because we use the stacked additive logistic normal distribution instead of mutually independent Dirichlet distributions. Meanwhile, $\boldsymbol{\Sigma}_b$'s off-diagonal terms govern the relationships within the top row of Table 1.

3.2. The observed data posterior

Thus far, we have not conditioned on the column totals; they are functions of the complete data, which are unobserved. To obtain the observed data posterior, we must sum out undetermined cells in each precinct table and integrate out the θs. Continuing with our Table 1 example, we sum out the four top left-hand cells, which produces

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{N}_{\mathrm{obs}}) \propto p(\boldsymbol{\mu}, \Sigma) \prod_{i=1}^{I} \left[\int \sum_{N_{bD_i} = lb_{N_{bD_i}}}^{ub_{N_{bD_i}}} \ \sum_{N_{bR_i} = lb_{N_{bR_i}}(N_{bD_i})}^{ub_{N_{bR_i}}(N_{bD_i})} \right. \tag{1}$$

$$\sum_{N_{wD_i} = lb_{N_{wD_i}}(N_{bD_i}, N_{bR_i})}^{ub_{N_{wD_i}}(N_{bD_i}, N_{bR_i})} \ \sum_{N_{wR_i} = lb_{N_{wR_i}}(N_{bD_i}, N_{bR_i}, N_{wD_i})}^{ub_{N_{wR_i}}(N_{bD_i}, N_{bR_i}, N_{wD_i})} \tag{2}$$

$$\binom{N_{b_i}}{N_{bD_i} \ N_{bR_i} \ N_{bA_i}} \binom{\mathrm{N}_{\mathrm{w_i}}}{N_{wD_i} \ N_{wR_i} \ N_{wA_i}} \binom{\mathrm{N}_{\mathrm{h_i}}}{N_{hD_i} \ N_{hR_i} \ N_{hA_i}} \tag{3}$$

$$\times \left(\theta_{bD_i}^{N_{bD_i}} \ \theta_{bR_i}^{N_{bR_i}} \ \theta_{bA_i}^{N_{bA_i}} \right) \left(\theta_{wD_i}^{N_{wD_i}} \ \theta_{wR_i}^{N_{wR_i}} \ \theta_{wA_i}^{N_{wA_i}} \right) \left(\theta_{hD_i}^{N_{hD_i}} \ \theta_{hR_i}^{N_{hD_i}} \ \theta_{hA_i}^{N_{hA_i}} \right) \tag{4}$$

$$\times |\Sigma|^{-1/2} \exp\{ -\tfrac{1}{2} (\boldsymbol{\omega}_i^* - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{\omega}_i^* - \boldsymbol{\mu}) \} \tag{5}$$

$$\times (\theta_{bD_i} \theta_{bR_i} \theta_{bA_i} \theta_{wD_i} \theta_{wR_i} \theta_{wA_i} \theta_{hD_i} \theta_{hR_i} \theta_{hA_i})^{-1} \tag{6}$$

$$\times I(N_{bD_i} + N_{wD_i} + N_{hD_i} = \mathrm{N}_{D_i}) \, I(N_{bR_i} + N_{wR_i} + N_{hR_i} = \mathrm{N}_{R_i}) \tag{7}$$

$$\times I(N_{bA_i} + N_{wA_i} + N_{hA_i} = \mathrm{N_{A_i}})\, I(N_{bD_i} + N_{bR_i} + N_{bA_i} = \mathrm{N_{b_i}}) \tag{8}$$

$$\times\, I(N_{wD_i} + N_{wR_i} + N_{wA_i} = \mathrm{N_{w_i}})\, I(N_{hD_i} + N_{hR_i} + N_{hA_i} = \mathrm{N_{h_i}}) \tag{9}$$

$$\times\, I(\theta_{bD_i} + \theta_{bR_i} + \theta_{bA_i} = 1)\, I(\theta_{wD_i} + \theta_{wR_i} + \theta_{wA_i} = 1) \tag{10}$$

$$\left. \times\, I(\theta_{hD_i} + \theta_{hR_i} + \theta_{hA_i} = 1)\, \mathrm{d}\boldsymbol{\theta}_i \right]. \tag{11}$$

$\mathbf{N}_{\mathrm{obs}}$ is a matrix, the ith row of which ($\mathbf{N}_{\mathrm{obs}_i}$) contains the row and column totals for precinct i's contingency table, and $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{b_i}^{\mathrm{T}}\ \boldsymbol{\theta}_{w_i}^{\mathrm{T}}\ \boldsymbol{\theta}_{h_i}^{\mathrm{T}})^{\mathrm{T}}$. 'lb' and 'ub' stand for lower bound and upper bound respectively. The notation in the summation symbols above reflects the fact that the ranges of the sum for inner summations depend on the values of the variable in outer summations. $\boldsymbol{\omega}_i^*$ is $\boldsymbol{\omega}_i$ thought of as a deterministic function of $\boldsymbol{\theta}_i$: not a random variable.

The above expression can be understood as follows. Lines (1) and (2) represent the prior as well as the integration over the missing Ns and θs. Lines (3) and (4) are the contribution from precinct i's multinomials: one for each precinct table row. Line (5) is the stacked additive logistic normal distribution. Line (6) is the Jacobian of the transformation from ω-space to θ-space. Lines (7)–(9) compel the rows and columns to sum to their respective observed totals. Lines (10) and (11) represent three separate sum-to-1 constraints for a precinct's θ-vector.

We fit the model by using a Gibbs sampler (Tanner, 1996), producing predictions of the internal cell counts (functions of which are often the quantities of interest) at every iteration in a manner that respects the bounds deterministically. Appendix A has the details.

3.3. Priors and quantities of interest

Identifying quantities of interest and assessing the implications of priors are especially critical in ecological inference, where the two are intimately connected. Wakefield (2004) demonstrated that, in this setting, seemingly innocuous choices of priors can have unexpected consequences, such as concentration of prior mass at extreme values of a parameter space. Moreover, we have found the ill-defined but near universal desire for 'flatness' in a prior a challenge to achieve. Priors that appear reasonably non-informative for one quantity of interest with one jurisdiction's racial pattern and for a particular election may no longer be so when applied to a different quantity of interest, or a different racial pattern or a different election.

For illustration, we discuss two sets of quantities of interest that are relevant to recent Texas redistricting litigation that reached the US Supreme Court (see LULAC *versus* Perry, 126 S. Ct. 2594 (2006) and Hebert *et al.* (2006)). As applied to 3 × 3 tables such as Table 1, the first consists of $\Lambda_{bD} = \Sigma_i N_{bD_i} / \Sigma_i\, (N_{bD_i} + N_{bR_i})$, and similarly for Λ_{wD} and Λ_{hD}, i.e. the fraction of each race's *voters* who support the Democrat. These quantities ordinarily determine whether voting is 'racially polarized'. The second set is the fraction of voters who are members of a particular race, i.e.

$$\Gamma_h = \sum_i (N_{hD_i} + N_{hR_i}) \Big/ \sum_i (N_{bD_i} + N_{bR_i} + N_{wD_i} + N_{wR_i} + N_{hD_i} + N_{hR_i})$$

for Hispanics. These quantities may show which race's voters control the election. Obviously, this list is not exhaustive; turnout by race ($\Sigma_i\, (N_{bD_i} + N_{bR_i}) / \Sigma_i\, \mathrm{N}_{b_i}$ for blacks) is often of interest.

Different quantities of interest may suggest different choices of prior parameters. To illustrate, we refer to Texas Congressional District 24, a district that played a prominent role in LULAC *versus* Perry. We use 2000 Presidential election results as matched to figures from the 2000 census by Lubin and Voss (2001). In 2000, the district had 249 precincts and a voting age population

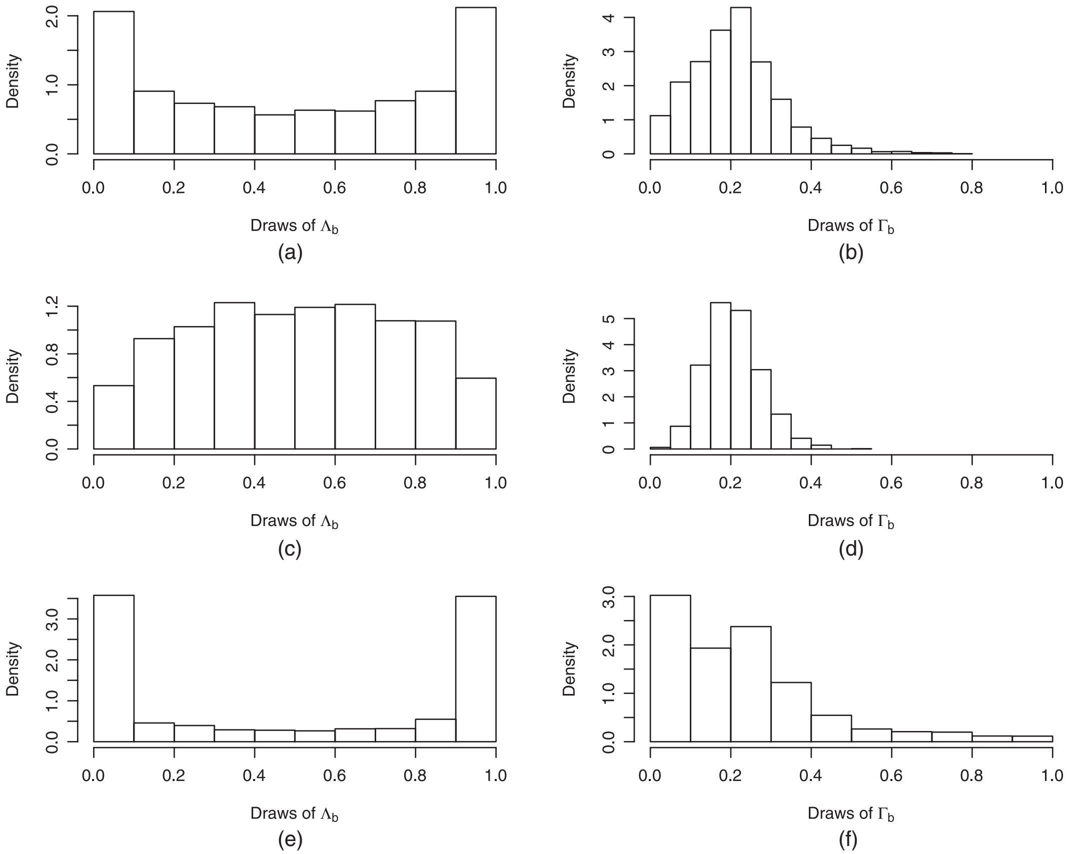

Fig. 1. Histograms of draws from the prior distribution of potential quantities of interest induced by different hyperparameter values (values for (a) and (b) are inappropriate for either quantity; values for (c) and (d) might be appropriate if the focus is on Λ_{bD}, but the nearly total lack of coverage at large values of Γ_b makes this prior potentially inappropriate for this quantity; values for (d) and (e) induce a bimodal (and thus unattractive) prior in Λ_{bD} but provide greater coverage for Γ_b across the (0,1) interval): (a), (b) $\nu_0 = 7$, $\psi_0 = 1.5$, $\kappa_0 = 5$, $\boldsymbol{\mu}_0 = \mathbf{-0.8}$; (c), (d) $\nu_0 = 10$; $\psi_0 = 1.5$, $\kappa_0 = 1$, $\boldsymbol{\mu}_0 = \mathbf{-0.8}$; (e), (f) $\nu_0 = 7$, $\psi_0 = 1$, $\kappa_0 = 25$, $\boldsymbol{\mu}_0 = \mathbf{-2.4}$

composition of 31% Hispanic, 20% non-Hispanic any part black and 49% other (white); there were six precincts that were 90% or more black, 14 that were 90% or more white and none that were 90% or more Hispanic.

Lacking a better idea, we set $\mathbf{K}_0 = \kappa_0 \mathbf{I}_{\dim(\boldsymbol{\Sigma})}$ and $\boldsymbol{\Psi}_0 = \psi_0 \mathbf{I}_{\dim(\boldsymbol{\Sigma})}$, where both κ_0 and ψ_0 are scalars. With this in mind, an instinctive approach to choosing prior values might suggest that ν_0 and ψ_0, which can be roughly conceptualized as the number of pseudodistricts and the precision that is added in the prior, should be set low, that κ_0 should be large (to reduce the influence of $\boldsymbol{\mu}_0$ on the posterior) and that $\boldsymbol{\mu}_0$ should be set under the expectation that approximately 50% of potential voters do not vote. The histograms in Figs 1(a) and 1(b) show simulations from a prior distribution that reflects such choices; in short, instincts are untrustworthy. The prior for Λ_{bD} is undesirable for an analysis focusing on this quantity because mass is concentrated at extreme values of parameter space, substantively corresponding to a belief that voting in the jurisdiction is polarized. Meanwhile, the prior that is induced on Γ_b is insufficiently diffuse, with values above 0.6 having essentially no mass. The histograms in Figs 1(c)–1(f) show potentially appropriate values for analyses in which interest is in Λ_{bD} (Figs 1(c) and 1(d)) and Γ_b (Figs 1(e) and 1(f)); note that hyperparameter values that induce a prior that is reasonable for one quantity might be less useful for the other quantity.

4. Simulation studies

We present the results of simulation studies of our model, which we conducted by using our running example of contingency tables of Table 1 form. Our research design begins with the identification of four dimensions to vary when generating data: the distribution of $(\mathrm{X}_{b_i}, \mathrm{X}_{w_i}, \mathrm{X}_{h_i})^{\mathrm{T}}$, which controls whether each table's units typically fall principally in one or another row or are spread more evenly across rows; κ_0, which controls the variation in expected behaviour of units across sets of contingency tables (data sets); $\boldsymbol{\mu}_0$, which controls whether (within each table) members of one row behave differently from members of another row in expectation; and the expected $\mathrm{N_i}$, corresponding to small and large numbers of units in each table. In terms of our running example, these dimensions translate to the level of housing segregation in a jurisdiction of interest, the expected uniformity of voting behaviour across jurisdictions, the expected extent of voting polarization and expected precinct size. These dimensions give us $2^4 = 16$ possible data-generating processes. We draw 100 data sets from each of 16 design combinations. Appendix A has the details.

We studied coverage rates, lengths of 90% and 95% posterior credible intervals (which were formed by equal-tailed quantiles) and other quantities with respect to the 21 hyperparameters in $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the nine internal cell count totals that are represented by $\Sigma_i N_{rc}$ for $r=$ (b, w, h) and $c=$ (D, R, A), Λ_{rD} for $r=$ (b, w, h) and turnout by race. We present some results for the quantity Λ_{hD} and for correlation coefficients of the $\boldsymbol{\Sigma}$-matrix. We focus on Λ_{hD} because, in our running example of US redistricting, inference about Hispanic voting patterns can be more important and more complex than inference for the corresponding whites and blacks patterns, the importance stemming from the potential for Hispanics to serve as swing voters, and the complexity stemming from their (generally) lower turnout rates (a circumstance that is mirrored in our simulations).

Coverage rates of all quantities for 90% and 95% intervals varied stochastically around nominal with no discernible pattern across simulation dimensions. For example, the rates for the 95% intervals of Λ_{hD} in the 16 simulations that are identified in Table 4 were (0.98, 0.97, 0.94, 0.95, 0.89, 0.94, 0.89, 0.94, 0.96, 0.98, 0.97, 0.93, 0.98, 0.94, 0.96, 0.96). More interesting was the variation in the lengths of the posterior intervals across simulation dimensions. Because the complete data (i.e. the internal cell counts) were drawn from the model, and because coverage rates were stochastically similar, these lengths measure the amount of information that the model can recover after the aggregation of internal cell counts to row and column margins.

Table 4 reports the (0.25, 0.75) quantiles of the lengths of 95% intervals for Λ_{hD} in the various simulation scenarios. Lower numbers mean smaller intervals and thus more information. A comparison of the top and bottom halves of Table 4 provides evidence of something that researchers have long guessed: that greater information is available when contingency table units tend to fall largely in one or another table row. More surprising is the fact that, at least for the (non-linear) function of the complete data represented by Λ_{hD}, greater information is available when the behaviour of units in row r is different from the behaviour of units in row r^*. The italic figures in Table 4 illustrate this fact: all else being equal, intervals for polarized data tend to be smaller than intervals for non-polarized data. Wilcoxon tests of a null hypothesis of equal mean interval lengths, run on the eight polar *versus* non-polar pairs of interval length vectors, resulted in p-values (to four decimal places) of (0, 0, 0, 0, 0, 0, 0.0061, 0.0017). Similarly, lower uniformity in expected voting behaviour tends to push values of $\boldsymbol{\mu}$ (across

Table 4. Quantiles of posterior interval lengths†

Segregation	*Polarization*	*Results for high uniformity*		*Results for low uniformity*	
		Large	*Small*	*Large*	*Small*
High	Polar	(0.049, 0.067)	(0.048, 0.065)	(0.041, 0.065)	(*0.040, 0.060*)
High	Non-polar	(0.060, 0.075)	(0.060, 0.079)	(0.054, 0.075)	(*0.053, 0.074*)
Low	Polar	(0.113, 0.145)	(0.118, 0.153)	(0.112, 0.163)	(0.103, 0.151)
Low	Non-polar	(0.136, 0.177)	(0.138, 0.181)	(0.126, 0.179)	(0.120, 0.162)

†Each cell represents 100 data sets drawn under different scenarios. 'High segregation' refers to highly segregated housing patterns; more generally, that each contingency table's units tend to concentrate within one row. 'High uniformity' refers to low variation in expected voter behaviour across jurisdictions, i.e. predictability of unit behaviour across data sets. 'Polar' refers to whether voting is racially polarized, i.e. whether (within each table) members of one row behave differently from members of another row (in expectation). 'Large' refers to the number of potential voters (units) in each precinct. The numbers in the cells represent the 0.25 and 0.75 quantiles of the lengths of the 95% posterior intervals for Λ_{hD}. Cell comparisons show how each simulation dimension affects the amount of information that is lost in aggregation. For example, comparison of the two cell values in italics suggests that, for highly segregated housing patterns in small precincts with low voter uniformity, greater information is available about Λ_{hD} when Hispanic voting patterns differ from those of blacks and whites (i.e. voting is polarized as opposed to non-polarized).

jurisdictions) into the extreme of the parameter space, making recovery of the bounded quantity Λ_{hD} easier.

Similar reasoning yields conclusions regarding the relative loss of information in between- and within-row relationships, a subject that previous $R \times C$ ecological inference models could not assess because the former were assumed away. Recall from Section 3.1 that, for example, the off-diagonal terms of $\boldsymbol{\Sigma}_b$ govern the relationships between the quantities within the top row of Table 1, whereas $\boldsymbol{\Sigma}_{bw}$ governs the relationships between the quantities in the top row of Table 1 *versus* those of the second row. To assess the comparative loss of information due to aggregation, we examined the coverage rates and posterior intervals for the three within-row correlations (one each for $\boldsymbol{\Sigma}_b$, $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}_h$) as compared with the 12 between-row correlations (four each for $\boldsymbol{\Sigma}_{bw}$, $\boldsymbol{\Sigma}_{bh}$ and $\boldsymbol{\Sigma}_{wh}$). Recovery rates of 95% intervals for both sets of quantities were nearly identical: 95.52% for within-row correlations *versus* 95.02% for between-row correlations (the difference was not statistically significant under standard tests for equal proportions). But, as Fig. 2 demonstrates, larger intervals were needed to recover between-row correlations. Wilcoxon tests suggested that these results were not due to random variation. Note also that Fig. 2 implies that, at least in some data sets, the data contain information about between- and within-row relationships in that the posterior intervals are sufficiently narrow for substantive conclusions (all this, of course, conditional on the model).

5. Application to real data: Bush *versus* Gore in Texas Congressional District 24, 2000

We apply our method to the Bush *versus* Gore Presidential contest as run in the precincts comprising Texas Congressional District 24 in 2000. One of the issues in LULAC *versus* Perry was whether Texas Congressional District 24 'performed' for African-Americans, despite the fact that blacks comprised only 20% of its voting age population. The allegation was that African-Americans could control the Democratic primary, then join with Hispanics and some whites

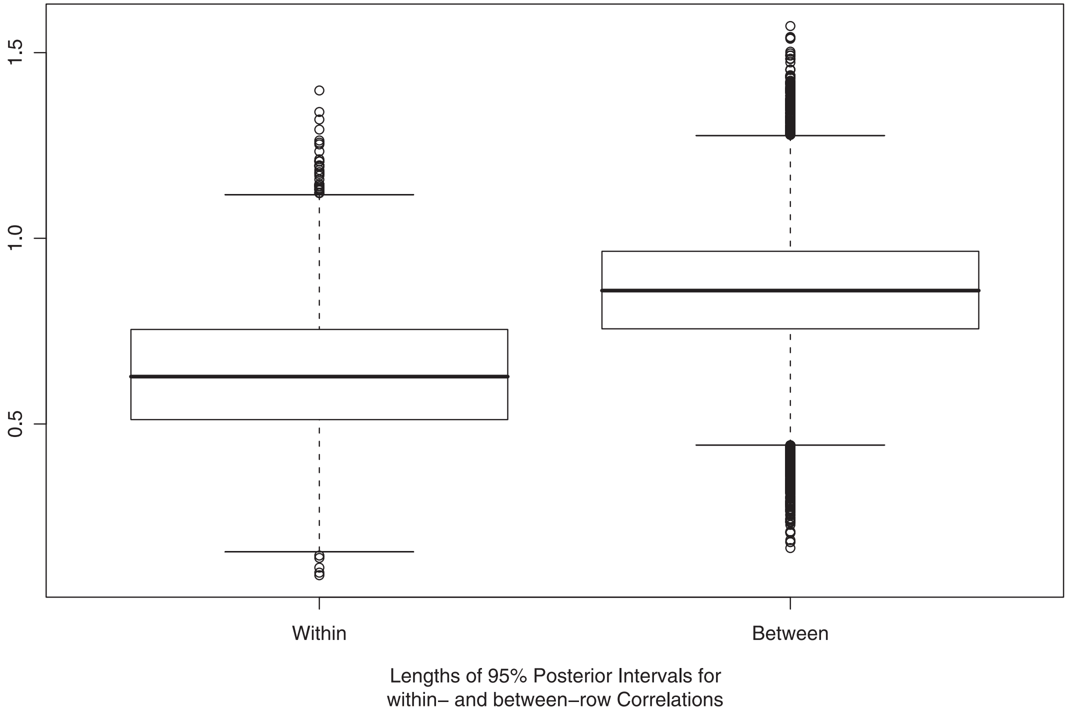

Fig. 2. Boxplots of the lengths of 95% posterior intervals for all between- and within-row correlations in the Σ-matrices from our simulations: the within-row correlations generally have smaller intervals than the between-row correlations, despite stochastically identical coverage rates; this suggests greater loss of information due to aggregation for between-row relationships

(the latter generally leaning Republican) to enable a Democratic victory. In 2000, the most competitive contest in the relevant precincts was the Presidential election, with Vice-President Gore winning 51–49; this contest also had the largest turnout.

The expert witnesses in LULAC *versus* Perry used a variant of the method of bounds as well as regression to assess racial voting patterns, but these methods yield either useless or physically impossible results for our data set. We fitted our model to the data by using five chains of 6 million iterations each, saving every 4000th draw. The method yielded 95% posterior intervals for Λ_{bD}, Λ_{wD} and Λ_{hD} as follows: (0.988–0.999), (0.224–0.262) and (0.955–0.999). 95% intervals for turnout for blacks, whites and Hispanics were (0.468–0.516), (0.445–0.469) and (0.079–0.113). In digesting these results, we proceed on two tracks: first, assessing their plausibility; second, drawing conclusions supposing the results to be accurate.

On the first point, all results seem quite reasonable except for the posterior distribution of Λ_{hD}, which suggests implausibly pro-Gore preferences among Hispanic voters. Informal consultations with people who were familiar with voting patterns in the area lent support to our view that the Λ_{hD}-results were suspicious. Fig. 3 demonstrates that one possible explanation for this result is aggregation bias. If white voters in predominantly Hispanic precincts supported Gore in greater fractions than white voters in predominantly white precincts, this fact, in combination with low turnout of Hispanics, might produce implausibly pro-Gore estimates for Hispanics. These results suggest that analysts should always compare a method's results with available external information for a data set. In our view, there are two primary defences against aggregation bias: obeying the bounds deterministically and adding information from

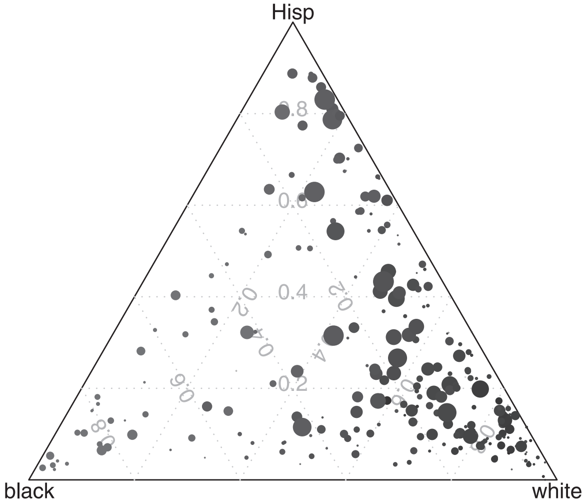

Fig. 3. Ternary plot of Bush *versus* Gore as run in the precincts constituting Texas Congressional District 24 in 2000: larger symbols indicate more populous precincts, and lighter shades indicate a greater Gore share of the two-party vote; if we drop an imaginary vertical line from the Hispanic vertex to the bisector of the triangle's bottom leg, we see that, generally, most non-Hispanics in predominantly Hispanic districts were white; if whites in predominantly Hispanic precincts voted more Democratic than whites in predominantly white precincts, aggregation bias could affect the estimates of the preferences of Hispanic voters, who turned out in low percentages

surveys, covariates or other sources. Our proposal builds in the former and facilitates the latter (see below). Of course, neither defence is foolproof, and it can be difficult to know whether one has implemented the latter properly (Cho, 1998).

Next, suppose that the results above were accurate. This is only one election, and there are various reasons why it might not be as indicative as other contests for inferences regarding racial bloc voting (e.g. both candidates are white). But if this pattern were evident in other general elections, and if blacks tended to dominate the Democratic primary (which is perhaps a reasonable hypothesis given their party preferences and turnout), then results like these support the assertion that District 24 performs for African-Americans, despite their 20% voting age population. African-American voters appear to be uniformly supporting a candidate of choice who is different from that of whites, but the black candidate of choice can prevail (barely) because of the support of the few Hispanics who vote and because of a small amount of crossover voting by whites.

6. Extensions

The ease with which our method may theoretically be extended is one of its appealing attributes. Extensions provide fertile ground for further research. We briefly mention a few here.

Because it begins with a model at the individual level, our method can incorporate certain kinds of survey information with relative ease. Continuing with our redistricting motivation, we confine our attention to a poll that might, with varying degrees of plausibility, be treated as a simple random sample of voters or potential voters in some precincts, perhaps a well-executed exit poll in a jurisdiction that, for legal reasons, tracks the races of people who enter polling booths (i.e. the Xs for each precinct table representing racial fractions of people entering or exiting polling places). We suppose that some form of polling is implemented in a subset S of the I precincts in the jurisdiction and contest of interest. In the ith precinct, let K_i denote the number polled, K_{b_i} the number of in-sample blacks and K_{bD_i} the number of in-sample blacks voting Democrat, with similar quantities defined for other races and parties. Recalling our individual level model, the likelihood of observing a particular vector $(\mathrm{K}_{bD_i}\ \mathrm{K}_{bR_i}\ \mathrm{K}_{bA_i}\ \mathrm{K}_{wD_i}\ \mathrm{K}_{wR_i}\ \mathrm{K}_{wA_i}\ \mathrm{K}_{hD_i}\ \mathrm{K}_{hR_i}\ \mathrm{K}_{hA_i})^{\mathrm{T}}$ in precinct i is

$$\left\{ \binom{\mathrm{N_i}}{\mathrm{K}_i}^{-1} \binom{\mathrm{K}_i}{\mathrm{K}_{b_i}\ \mathrm{K}_{w_i}\ \mathrm{K}_{h_i}} \mathrm{X}_{b_i}^{\mathrm{K_{b_i}}} \mathrm{X}_{w_i}^{\mathrm{K_{w_i}}} \mathrm{X}_{h_i}^{\mathrm{K_{h_i}}} \right. \tag{12}$$

$$\times \binom{\mathrm{K}_{b_i}}{\mathrm{K}_{bD_i}\ \mathrm{K}_{bR_i}\ \mathrm{K}_{bA_i}} \theta_{bD_i}^{\mathrm{K}_{bD_i}} \theta_{bR_i}^{\mathrm{K}_{bR_i}} \theta_{bA_i}^{\mathrm{K}_{bA_i}} \tag{13}$$

$$\times \binom{\mathrm{K}_{w_i}}{\mathrm{K}_{wD_i}\ \mathrm{K}_{wR_i}\ \mathrm{K}_{wA_i}} \theta_{wD_i}^{\mathrm{K}_{wD_i}} \theta_{wR_i}^{\mathrm{K}_{wR_i}} \theta_{wA_i}^{\mathrm{K}_{wA_i}} \tag{14}$$

$$\left. \times \binom{\mathrm{K}_{h_i}}{\mathrm{K}_{hD_i}\ \mathrm{K}_{hR_i}\ \mathrm{K}_{hA_i}} \theta_{hD_i}^{\mathrm{K}_{hD_i}} \theta_{hR_i}^{\mathrm{K}_{hR_i}} \theta_{hA_i}^{\mathrm{K}_{hA_i}} \right\}^{I(i \in S)}. \tag{15}$$

In the posterior, after conditioning on the observed Ks (and thus adjusting, for each $i \in S$, the applicable Ns), adding lines (12)–(15) to lines (1)–(11) does not disrupt the basic nature of the model. In particular, the fitting process that is articulated in Appendix A for the basic model can proceed largely as before.

Matters become more complicated if the jurisdiction does not record the races of people entering polling booths (i.e. the Xs for each precinct table representing racial fractions of people who could enter or exit polling places). Even accepting that we observe a simple random sample in precinct i of people exiting the polling place, it may be implausible to assume that the frame corresponding to this sample is the set of all potential voters in the precinct because too many potential voters stay away from the polls. One way to proceed here is to discard the information from any survey respondent who reports that he or she did not vote in the electoral contest of interest and to assume that the responses of those who did vote (again totalling K_i in our notation) constitute a simple random sample in precinct i of actual voters. Under this assumption, the expression corresponding to lines (12)–(15) for the previously discussed case is

$$\left\{ \binom{\mathrm{N_i} - \mathrm{N_{A_i}}}{\mathrm{K}_i}^{-1} \binom{\mathrm{K}_i}{\mathrm{K}_{b_i}\ \mathrm{K}_{w_i}\ \mathrm{K}_{h_i}} \left(\frac{\mathrm{N_{b_i}} - N_{bA_i}}{\mathrm{N_i} - \mathrm{N_{A_i}}} \right)^{\mathrm{K}_{b_i}} \left(\frac{\mathrm{N_{w_i}} - N_{wA_i}}{\mathrm{N_i} - \mathrm{N_{A_i}}} \right)^{\mathrm{K}_{w_i}} \left(\frac{\mathrm{N_{h_i}} - N_{hA_i}}{\mathrm{N_i} - \mathrm{N_{A_i}}} \right)^{\mathrm{K}_{h_i}} \right. \tag{16}$$

$$\times \binom{\mathrm{K}_{b_i}}{\mathrm{K}_{bD_i}\ \mathrm{K}_{bR_i}} \left(\frac{\theta_{bD_i}}{1 - \theta_{bA_i}} \right)^{\mathrm{K}_{bD_i}} \left(\frac{\theta_{bR_i}}{1 - \theta_{bA_i}} \right)^{\mathrm{K}_{bR_i}} \tag{17}$$

$$\times \binom{\mathrm{K}_{w_i}}{\mathrm{K}_{wD_i}\ \mathrm{K}_{wR_i}} \left(\frac{\theta_{wD_i}}{1 - \theta_{wA_i}} \right)^{\mathrm{K}_{wD_i}} \left(\frac{\theta_{wR_i}}{1 - \theta_{wA_i}} \right)^{\mathrm{K}_{wR_i}} \tag{18}$$

$$\left. \times \binom{\mathrm{K}_{h_i}}{\mathrm{K}_{hD_i}\ \mathrm{K}_{hR_i}} \left(\frac{\theta_{hD_i}}{1 - \theta_{hA_i}} \right)^{\mathrm{K}_{hD_i}} \left(\frac{\theta_{hR_i}}{1 - \theta_{hA_i}} \right)^{\mathrm{K}_{hR_i}} \right\}^{I(i \in S)}. \tag{19}$$

In terms of sampling from the posterior, matters have become less agreeable. Any term involving a θ or an italicized N cannot be ignored, and the presence of such terms in line (16) means that the fitting strategy that is articulated in Appendix A is no longer directly applicable. Outside the exit polling context, other sampling schemes (e.g. stratified or cluster sampling) raise additional challenges, although addressing such challenges may be eased by the presence of a coherent individual level model.

Other, non-sampling, extensions are also possible. For instance, in some applications it may be useful to allow row- or column-specific random effects to influence $\boldsymbol{\omega}_i$. These effects could be assumed to be independent across tables or could be parameterized to allow for spatial, temporal or spatiotemporal associations across tables (Haneuse and Wakefield, 2004; Quinn, 2004). The utility of such modelling strategies will obviously vary greatly across applications.

Finally, our model is amenable to extensions that have been previously discussed in the literature. Covariates may be included by substituting regressions for $\boldsymbol{\mu}$. Robustness might be increased via the substitution of a more dispersed ellipsoidally symmetric distribution, such as multivariate t with known or unknown degrees of freedom, for the normal distribution. The normal structure at the upper level facilitates sharing of information between electoral contests across time and office via hierarchical techniques.

We caution that in applied work the analyst must consider whether possible extensions will be appropriate. The loss of information due to aggregation is known to be large, and adding complications to the basic method might demand too much from the data. Our point here is that our method provides a flexible structure within which researchers can explore what the data can tolerate and, correspondingly, how much can be learned.

7. Conclusion

In this paper, we have proposed a count ecological inference model that can handle data sets with precinct tables of any size and shape. Our method has certain advantages over previously proposed models, such as deterministic respect for the bounds, flexibility and correspondence to a plausible account of the data-generating process. We have discussed factors affecting the choice of prior in the redistricting context and demonstrated that careful thought is required here. We have used simulation studies from our model to assess the relative losses of information due to aggregation for various quantities that might be of interest to applied researchers. In substantive applications, we have shown how the results of our model may be useful in redistricting litigation.

Acknowledgements

The authors extend great thanks to Aaron Culich and Jayanta Sircar at the Harvard School of Engineering, and to the stewards of the Research Computing Environment at the Harvard–Massachusetts Institute of Technology Data Center, for their assistance in providing computing power. The authors also thank Brady Baybeck, Andrew Gelman, Jeff Gill, Adam Glynn, Daniel Ho, Kosuke Imai, George Judge, Gary King, Jon Wakefield, a referee and the Associate Editor for helpful comments on an earlier version of this paper. In addition, Quinn thanks the US National Science Foundation (grant SES 03-50613) and the Center for Advanced Study in the Behavioral Sciences for its hospitality and support. Any errors are those of the authors alone.

Appendix A

A.1. *Fitting*

To fit our model, we create a Gibbs framework (Tanner, 1996) in which we successively sample from

(a) the distribution of the internal cell counts in each precinct given $\boldsymbol{\theta}_i$ in a manner that respects the bounds deterministically,

(b) the distribution of each precinct's $\boldsymbol{\theta}_i$ given that precinct's internal cell counts as well as $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and
(c) the distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given all precincts' $\boldsymbol{\theta}_i$s.

Regarding the internal cell counts, $p(N_{bD_i}, N_{bR_i}, N_{bA_i}, N_{wD_i}, N_{wR_i}, N_{wA_i}, N_{hD_i}, N_{hR_i}, N_{hA_i} | \mathbf{N}_{\text{obs}_i}, \boldsymbol{\theta}_i)$ is proportional to the product of lines (3) and (4) and (7)–(9) in Section 3.2. To draw samples from this conditional posterior, we adapt the algorithm that was proposed by McDonald *et al.* (1999) for sampling from a roughly similar distribution. The key point is that the distribution of the counts in a 2×2 precinct subtable that is defined by the intersection of any two unique rows and any two unique columns, conditional on all remaining internal counts, is univariate non-central hypergeometric. This fact suggests a small Gibbs sampling strategy in which we choose two rows and two columns in the precinct table, form the corresponding 2×2 subtable together with the current values of its four θ-parameters, add to calculate the row and column totals of the 2×2 subtables, normalize the θs, discard the original counts in the internal cells of the subtable and draw from the corresponding univariate non-central hypergeometric distribution by using an adaptation of the method that was developed by Liao and Rosen (2001). We complete this small Gibbs step at least once for each possible combination of two rows and two columns within each precinct within each iteration of the overall Gibbs sampling.

Regarding the row probabilities, in each precinct, the distribution of $\boldsymbol{\theta}_i$ given all the internal cell counts as well as $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is proportional to the product of lines (4)–(6) and (10) and (11), above. Because this distribution is non-standard, we sample from it by using the Metropolis–Hastings algorithm (Tanner, 1996). We generate a proposal in $\boldsymbol{\omega}$-space from a multivariate $t_4(\boldsymbol{\mu}^{(t)}, \gamma_i \boldsymbol{\Sigma}^{(t)})$ distribution and transform back to $\boldsymbol{\theta}$. γ_i is a (constant) tuning parameter which is set in initial runs, and superscript (t) denotes the iteration. The conditional distributions of the hyperparameters are in standard form.

A.2. Simulation studies

We began by drawing $(\mathrm{X}_{b_i}, \mathrm{X}_{w_i}, \mathrm{X}_{h_i}) \sim \mathrm{Diri}\{\alpha * (0.35, 0.45, 0.2)\}$, a process that in expectation yields population shares of African-Americans, whites and Hispanics that are similar to those in jurisdictions of frequent interest. $\alpha = 0.5$ and $\alpha = 2.5$ corresponded to high and low segregation in housing respectively. We drew $\mathrm{N}_i \sim \mathrm{Poi}(Q)$, with $Q = 400$ and $Q = 2000$ corresponding to small and large precincts respectively. We then drew $\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \kappa_0 \mathbf{I}_6)$. For polarized and non-polarized simulations, $\boldsymbol{\mu}_0 = (0, -2, -2, 0, -0.5, -1.5)$ and $\boldsymbol{\mu}_0 = (-0.7, -0.7, -0.45, -0.45, -0.74, -0.74)$ respectively, mirroring (in expectation) party support levels and turnout rates that are often found in voting patterns by race. For high and low uniformity, we set $\kappa_0 = 0.05$ and $\kappa = 0.5$ respectively. Replication information appears on our Web sites `http://tinyurl.com/2ynrwm` and `http://tinyurl.com/32kumh`.

We assessed convergence by examining diagnostics that were proposed by Gelman and Rubin (1992) and Heidelberger and Welch (1983) and by reviewing auto-correlation measures. Stubborn auto-correlation necessitated the use of 1 500 000 iterations per chain, and we ran three chains for each simulation, discarding the first 15% of each as a burn-in. These settings resulted in generally unremarkable values for the convergence diagnostics. For example, for the 27 upper level parameters in our multivariate normal distribution, over 97% of the Gelman–Rubin diagnostics had values below 1.1, and fewer than 2.5% had Heidelberger–Welch p-values of 0.05 or lower.

References

Achen, C. H. and Shively, W. P. (1995) *Cross-level Inference*. Chicago: University of Chicago Press.

Aitchison, J. (2003) *The Statistical Analysis of Compositional Data*, 2nd edn. Caldwell: Blackburn.

Benoit, K., Laver, M. and Giannetti, D. (2004) Multiparty split-ticket voting estimation as an ecological inference problem. In *Ecological Inference: New Methodological Strategies* (eds G. King, O. Rosen and M. A. Tanner). Cambridge: Cambridge University Press.

Brown, P. J. and Payne, C. D. (1986) Aggregate data, ecological regression, and voting transitions. *J. Am. Statist. Ass.*, **81**, 452–460.

Cho, W. K. T. (1998) Iff the assumption fits . . . : a comment on the King ecological inference solution. *Polit. Anal.*, **7**, 143–163.

Duncan, O. D. and Davis, B. (1953) An alternative to ecological correlation. *Am. Sociol. Rev.*, **18**, 665–666.

Ferree, K. E. (2004) Iterative approaches to R × C ecological inference problems: where they can go wrong and one quick fix. *Polit. Anal.*, **12**, 143–159.

Freedman, D. A., Klein, S. P., Sacks, J., Smyth, C. A. and Everett, C. G. (1991) Ecological regression and voting rights. *Evaln Rev.*, **15**, 673–711.

Gelman, A., Park, D. K., Ansolabehere, S., Price, P. N. and Minnite, L. C. (2001) Models assumptions and model checking in ecological regressions. *J. R. Statist. Soc.* A, **164**, 101–118.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.*, **7**, 457–511.

Greiner, D. J. (2007) Ecological inference in Voting Rights Act disputes. *Jurmetr. J.*, **47**, 115–167.

Grofman, B., Migalski, M. and Noviello, N. (1985) The "Totality of the circumstances test" in Section 2 of the 1982 Extension of the Voting Rights Act: a social science perspective. *Law Poly*, **7**, 199–223.

Haneuse, S. and Wakefield, J. (2004) Ecological inference incorporating spatial dependence. In *Ecological Inference: New Methodological Strategies* (eds G. King, O. Rosen and M. A. Tanner). Cambridge: Cambridge University Press.

Hebert, J. G., Smith, P. M. and Hirsch, S. (2006) Brief for Appellants in Jackson v. Perry, 05-276. US Supreme Court.

Heidelberger, P. and Welch, P. D. (1983) Simulation run length control in the presence of an initial transient. *Ops Res.*, **31**, 1109–1144.

Judge, G., Miller, D. J. and Cho, W. K. T. (2004) An information theoretic approach to ecological estimation and inference. In *Ecological Inference: New Methodological Strategies* (eds G. King, O. Rosen and M. A. Tanner). Cambridge: Cambridge University Press.

King, G. (1997) *A Solution to the Ecological Inference Problem*. Princeton: Princeton University Press.

Liao, J. G. and Rosen, O. (2001) Fast and stable algorithms for computing and sampling from the noncentral hypergeometric distribution. *Am. Statistn*, **55**, 366–369.

Lublin, D. and Voss, D. S. (2001) *Federal Elections Project*. American University, Washington DC. (Available from `http://spa.american.edu/ccps/pages.php?ID=12`.)

McDonald, J. W., Smith, P. W. F. and Forster, J. J. (1999) Exact tests of goodness of fit of log-linear models for rates. *Biometrics*, **55**, 620–624.

Prentice, R. L. and Sheppard, L. (1995) Aggregate data studies of disease risk factors. *Biometrika*, **82**, 113–125.

Quinn, K. M. (2004) Ecological inference in the presence of temporal dependence. In *Ecological Inference: New Methodological Strategies* (eds G. King, O. Rosen and M. A. Tanner). Cambridge: Cambridge University Press.

Robinson, W. S. (1950) Ecological correlations and the behavior of individuals. *Am. Sociol. Rev.*, **15**, 351–357.

Rosen, O., Jiang, W., King, G. and Tanner, M. A. (2001) Bayesian and frequentist inference for ecological inference: the R × C case. *Statist. Neerland.*, **55**, 134–156.

Tanner, M. A. (1996) *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 3rd edn. New York: Springer.

Wakefield, J. (2004) Ecological inference for 2 × 2 tables (with discussion). *J. R. Statist. Soc.* A, **167**, 385–445.