# PART TWO

# 4 Extending King's Ecological Inference Model to Multiple Elections Using Markov Chain Monte Carlo

Jeffrey B. Lewis

ABSTRACT

King's EI estimator has become a widely used procedure for tackling so-called ecological inference problems. The canonical ecological inference problem involves inferring the rate of voter turnout among two racial groups in a set of electoral precincts from observations on the racial composition and total voter turnout in each precinct. As a Bayesian hierarchical model, EI links information about the turnout by race in each precinct to information turnout by race in other precincts through the assumption that turnout rates are independently drawn from a common distribution. In this way, strength is borrowed from other precincts in estimating the turnout rates by race within each precinct. Commonly, marginal turnout rates and racial compositions are observed for multiple elections within the same set of aggregate units. This chapter extends King's estimator to this case, allowing strength to be borrowed not only across precincts within the same election, but also across elections within precincts. The model is estimated via an MCMC algorithm, validated using simulated data, and applied to estimating voter turnout by race in Virginia during the 1980s.

## 4.1 INTRODUCTION

King's (1997) EI estimator has become a widely used procedure for tackling so-called ecological inference problems. The canonical ecological inference problem involves inferring the rate of voter turnout among two racial groups in a set of electoral precincts from observations on the racial composition and total voter turnout in each precinct.[1] As a Bayesian hierarchical model, EI links information about the turnout by race in each precinct to information turnout by race in other precincts through the assumption that turnout rates are independently drawn from a common distribution. In this way, strength is borrowed from all precincts in estimating the turnout rates by race within each precinct. Commonly, marginal turnout rates and racial compositions are observed for multiple elections within the same set of precincts. This chapter extends King's estimator to this case, allowing strength to be borrowed not only across precincts within the same election, but also across elections within precincts.

One common use of EI is estimating turnout by race or estimating support for particular candidates by race. For example, EI has been used to assess claims about "racial block" voting that are often central to court cases involving legislative redistricting. The central question is whether voters of a particular racial or ethnic group in a state or locality habitually vote the same candidates. In these cases, election returns from various contests and across a number

[1] Here "precincts" may be electoral precincts or any other aggregate groupings of votes for which the racial composition is known.

of election cycles are typically available. While it is reasonable to assume that turnout rates or voting patterns in a given set of precincts are correlated across elections or contests, this additional source of information has not been exploited by previous estimators (one exception is Quinn, in this book, Chapter 9).

Given the difficulty of ecological inference and particularly the difficulty of estimating precinct quantities of interest, building a model that capitalizes on the commonalities that might exist across contests or elections at the precinct level is potentially fruitful. As demonstrated below, by borrowing strength across elections as well as across precincts, the mean square error of the precinct level predictions can be substantially improved. The model builds directly on King's EI. Turnout rates among whites and nonwhites across precincts are assumed to be drawn from a truncated bivariate normal distribution (TBVN). Whereas King's EI models the parameters of the truncated bivariate normal as election-specific, in the multielection model the parameters of the truncated bivariate normal distribution are determined by both precinct-specific and election specific effects. In this way, the estimation of turnout rates by racial group are tied not only across precincts within elections as in King's model, but also across elections within precincts. The resulting seemingly unrelated ecological inference (SUEI) model is presented in detail below.[2]

The increased complexity introduced by the precinct-specific determinants of the underlying truncated bivariate normal that describes the distribution of the precinct-level turnout by race makes estimation by the standard maximization approach employed by King infeasible. Instead, the model is estimated by Markov chain Monte Carlo (MCMC). King's EI and the SUEI are members of a general class of hierarchical and mixture models that are known to be amenable to estimation by MCMC methods.

The estimator is applied to sets of simulated data and to electoral data from Virginia. Because the "true" values of the precinct quantities are known for the simulated data, the advantage of SUEI over King's EI can be directly assessed. For the Virginia data, the true values are not known. However, the estimates generated by each method can still be compared and the strength of the cross-election precinct effects assessed. Overall, SUEI and EI yield quite similar estimates of the aggregate quantities of interest, but in some cases the two models find quite different estimates of the precinct quantities. By borrowing strength across elections, SUEI is able to reduced the mean square error (MSE) of the precinct-level estimates in the simulated data by as much 40 to 50 percent. In the Virginia data sizeable cross-election dependencies are found and precinct-level estimates differ substantially from EI to SUEI.

## 4.2 KING'S EI MODEL

As presented in King (1997) and extensively discussed elsewhere, the basic EI model has as its foundation an identity, an independence assumption, and a distributional assumption that together form a simple mixture model. The identity says that in each precinct the total turnout rate must be the sum of the fraction of nonwhites that vote and the fraction of whites that vote weighted by the fraction that each group comprises in the voting-age population of the precinct (see Introduction, Equation 4). Formally, let $\mathbf{T} = (T_1, T_2, \ldots, T_p)$ be the voter turnout rates in a set of $p$ precincts, and $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ be the fractions of the

[2] The notion of "seemingly unrelated" EI follows by analogy from the seemingly unrelated regression (Zellner, 1962) model in which linear regression models are connected only through correlations among their stochastic components. Here EI models that would otherwise be estimated separately are linked through a common precinct-specific stochastic component.

populations in each of the same set of precincts that is nonwhite. Then

$$T_i = \beta_i^b X_i + \beta_i^w (1 - X_i) \qquad \text{for } i = 1, \ldots, p, \tag{4.1}$$

where $\boldsymbol{\beta}^b = (\beta_1^b, \beta_2^b, \ldots, \beta_p^b)$ and $\boldsymbol{\beta}^w = (\beta_1^w, \beta_2^w, \ldots, \beta_p^w)$ are unobserved turnout rates among blacks and whites in the precincts. **T** and **X** are known from election returns and census data, while $\boldsymbol{\beta}^b$ and $\boldsymbol{\beta}^w$ are unknown quantities to be estimated. Because there are twice as many unknown quantities to be estimated as observations, additional assumptions must be made to identify the model. The $(\beta^b, \beta^w)$ pairs for each precinct are assumed to be drawn independently from a common joint density. In particular, the $(\beta^b, \beta^w)$ pairs are assumed to be drawn from a truncated bivariate normal distribution with parameters $\breve{\psi} = (\breve{\mathfrak{B}}^b, \breve{\mathfrak{B}}^w, \breve{\sigma}_b, \breve{\sigma}_w, \breve{\rho})$.[3] The truncation is on the unit square, reflecting the logical bounds of $\beta^b$ and $\beta^w$, which, as fractions of populations, must fall between zero and one.

Suppressing the precinct subscripts and noting that $T$ is a linear function of the random quantities $\beta^b$ and $\beta^w$, standard change-of-variables techniques yield the joint distribution of $\beta^b$ and $T$.[4] As shown in Lewis (2002), the joint distribution of $\beta^b$ and $T$ is bivariate truncated normal with region of support $\{(\beta^b, T) \text{ s.t. } 0 \leq \beta^b \leq 1 \;\&\; \beta^b X \leq T \leq \beta^b X + (1 - X)\}$. Given that $\beta^b$ and $T$ are truncated bivariate normal, it is easy to show that $\beta^b | T$ is truncated normal (see King, 1997, or Lewis, 2002). Let $f$ be the joint density of $\beta^b$ and $\beta^w$, and $g$ be the joint density of $\beta^b$ and $T$. For simplicity, I will parameterize $g$ by $\breve{\psi}$ and $X$.[5]

In order to estimate the posterior distribution of $\breve{\psi}$, King marginalizes the joint distribution of $\beta^b$ and $T$ with respect to $\beta^b$ to find

$$g(T | X, \breve{\psi}) = \int_0^1 g(\beta^b, T | X, \breve{\psi}) d\beta^b.$$

Given the assumption of independent sampling, the likelihood of the observed data can be written as

$$L(\mathbf{T} | \mathbf{X}, \breve{\psi}) = \prod_i g(T_i | X_i, \breve{\psi}).$$

3 This notation matches King (1997). King considers both the expectations and covariances of the distribution of $\beta^b$ and $\beta^w$ and the parameters of the truncated bivariate normal which describes the means and variances of the "corresponding untruncated variables" (p. 102). The later quantities are what constitute $\breve{\psi}$. Also, see Kings's introduction (Equation 6 and surrounding text).

4 $X$ is taken to be a fixed quantity.

5 The joint distribution of $T$ and $\beta^b$ is

$$g(\beta^b, T; \breve{\psi}) = \frac{\phi_2(\beta^b, T; M(\breve{\psi}))}{\int_0^1 \int_0^1 \phi_2(\beta^b, \beta^w; \breve{\psi}) d\beta^b d\beta^w},$$

where $\phi_2$ is the bivariate normal density function and $M$ transforms the parameters of the joint distribution of $\beta^b$ and $\beta^w$ into the parameters of the joint distribution of $\beta^b$ and $T$.

$$M : \left(\breve{\mathfrak{B}}^b, \breve{\mathfrak{B}}^w, \breve{\sigma}_b, \breve{\sigma}_w, \breve{\rho}\right) \longrightarrow \left(\breve{\mathfrak{B}}^b, \breve{\mathfrak{B}}^b X + \breve{\mathfrak{B}}^w (1 - X), \breve{\sigma}_b, \right.$$
$$\left. \sqrt{\breve{\sigma}_b^2 X^2 + \breve{\sigma}_w^2 (1 - X)^2 + 2\breve{\rho}\breve{\sigma}_b\breve{\sigma}_w X(1 - X)}, \frac{\breve{\sigma}_b X + \breve{\sigma}_w \rho (1 - X)}{\sqrt{\breve{\sigma}_b^2 X^2 + \breve{\sigma}_w^2 (1 - X)^2 + 2\breve{\rho}\breve{\sigma}_b\breve{\sigma}_w X(1 - X)}}\right).$$

These expressions hold for $0 \leq X < 1$. If $X = 1$, then $T = \beta^b$ and the joint distribution of $T$ and $\beta^b$ is simply the marginal distribution of $\beta^b$. In what follows, I avoid this technical nuisance by replacing $X$ with $X - \epsilon$ in the data if $X = 1$.

Formulas for $L$ are given in King (1997, Appendix D) and are not repeated here. The posterior distribution of $\breve{\psi}$ given the data is

$$\mathrm{P}(\breve{\psi}|\mathbf{T},\mathbf{X}) \propto L(\mathbf{T}|\mathbf{X},\breve{\psi})\mathrm{p}(\breve{\psi}), \tag{4.2}$$

where p is the joint prior distribution over $\breve{\psi}$.

Default options for King's computer implementation of EI place flat (improper) priors over $\breve{\mathfrak{B}}^b$ and $\breve{\mathfrak{B}}^w$, diffuse half-normal priors over $\breve{\sigma}^b$ and $\breve{\sigma}^w$, and an informative normal prior over the Fischer's $Z$ transformation of $\breve{\rho}$. The informative prior on $\breve{\rho}$ effectively bounds estimates of its posterior mode away from 1 and −1. As noted by King, there is little information in the data about $\breve{\rho}$, and at extreme values of $\breve{\rho}$ the calculation of $L$ becomes unreliable.[6] The prior distributions of each element of $\breve{\psi}$ are taken to be independent.

King's estimates the posterior distribution of $\breve{\psi}$ using numerical maximization of P to find the posterior modes and then uses normal asymptotic theory augmented with importance resampling to simulate draws from P.[7]

Given the posterior distribution of $\breve{\psi}$, the posterior distribution of each $\beta_i^b$ (or $\beta_i^w$) given **T** and **X** can be formed as

$$\mathrm{P}\big(\beta_i^b|\mathbf{T},\mathbf{X}\big) = \int g\big(\beta_i^b|T_i, X_i, \breve{\psi}\big)\mathrm{P}\big(\breve{\psi}|\mathbf{T},\mathbf{X}\big)\, d\breve{\psi}. \tag{4.3}$$

While the integral in Equation 4.3 is difficult to evaluate directly, it is easy to draw samples from this density using Gibbs sampling. Suppressing the **X** and **T** from the notation, $\breve{\psi}^*$ is drawn from $\mathrm{P}(\breve{\psi})$ using asymptotic normality and importance resampling, and then a draw is made from $g(\beta_i^b|T_i, X_i, \breve{\psi}^*)$ conditional on $\breve{\psi}^*$. Because $\beta^b|T$ is distributed truncated normal, $g(\beta_i^b|T_i, X_i, \breve{\psi}^*)$ can be sampled from using inverse CDF sampling. Samples from the posterior distribution of $\beta_i^b|T_i$ made in this way can be used to draw histograms or to calculate a posteriori expectations of these precinct-level quantities of interest.

Equation 4.3 reveals that through the assumption that all $\beta^b$ and $\beta^w$ pairs are drawn from a common distribution, EI "borrows strength" from data for other precincts in estimating the value of $\beta^b$ in each precinct even though the draws for each precinct are a priori independent. By the accounting identity, given $T_i$, $\beta^b$ and $\beta^w$ are linearly dependent:[8]

$$\beta_i^w = \frac{T_i}{1-X_i} - \frac{X_i}{1-X_i}\beta_i^b.$$

Thus, the posterior distribution of $\beta_i^w$ can be estimated using samples drawn from the distribution of $\beta_i^b$. King uses samples from the posterior distribution of the precinct $\beta$'s to calculate other quantities of interest, such as the election-wide rates of turnout among blacks and whites.

### 4.2.1 MCMC Estimation of EI

As an alternative to King's procedure, I have implemented an MCMC estimator for the probability model described above. The estimator has the typical advantages of MCMC

[6] $L$ requires the calculation of the bivariate normal integral over the unit square. This calculation can become noisy at extreme values of $\breve{\rho}$ or, more generally, whenever the area over the unit square is very small.

[7] In order to improve the normal approximation to the posterior distribution and to decrease the posterior correlations among the parameters, King first reparameterizes the posterior distribution, maximizing over the logs of $\breve{\sigma}_b$ and $\breve{\sigma}_w$, the Fischer's $z$ transformation of $\breve{\rho}$, and $(\breve{\mathfrak{B}}^b - 0.5)/(0.25 + \breve{\sigma}_b)$ and $(\breve{\mathfrak{B}}^w - 0.5)/(0.25 + \breve{\sigma}_w)$ rather than $\breve{\mathfrak{B}}^b$ and $\breve{\mathfrak{B}}^w$.

[8] See (*This volume*) Introduction, p. 4, Equation 5 and surrounding text.

over King's procedure: it yields draws from the exact posterior distributions and is (in principle) more robust to numerical inaccuracies. It also has the typical disadvantages: lack of speed, difficulty in determining convergence, and so forth.

Rather than marginalizing $\boldsymbol{\beta}^b$ in forming the posterior distribution of $\breve{\boldsymbol{\psi}}$ as described above, in the MCMC approach the complete joint posterior distribution of $(\breve{\boldsymbol{\psi}}, \boldsymbol{\beta}^b)$ is recovered. In this way, the joint distribution of the precinct quantities of interest are obtained directly from the estimation. Implementing MCMC using Gibbs sampling is straightforward. First, provisional values for $\breve{\boldsymbol{\psi}}$ are set. Next, values for $\beta_i^b$ conditional on the provisional $\breve{\boldsymbol{\psi}}$, $T_i$, and $X_i$ are drawn for $i = 1, 2, \ldots, p$. Then values of each element of $\breve{\boldsymbol{\psi}}$ are drawn conditional on the sampled values of the $\boldsymbol{\beta}^b$, the values of the other elements of $\breve{\boldsymbol{\psi}}$, and the data. This process is then repeated. In the limit, the distribution of the sampled values will follow the joint posterior distribution of $\boldsymbol{\beta}^b$ and $\breve{\boldsymbol{\psi}}$ (see Gamerman, 1997).

As noted above, $g(\beta^b | T, X, \breve{\boldsymbol{\psi}})$ is truncated normal and can be sampled from using inverse CDF sampling. The more difficult distribution from which to sample is

$$\mathrm{P}\left(\breve{\boldsymbol{\psi}} | \boldsymbol{\beta}^b, \mathbf{T}, \mathbf{X}\right) \propto \left(\prod_i g\left(T_i, \beta_i^b | \breve{\boldsymbol{\psi}}\right)\right) \mathrm{p}(\breve{\boldsymbol{\psi}}).$$

I use adaptive rejection Metropolis sampling (ARMS; Gilks, Best, and Tan, 1995) to draw from the conditional distribution of each element of $\breve{\boldsymbol{\psi}}$ conditional on the prior values of the other elements.[9] As shown by Gilks et al., ARMS allows sampling from arbitrary distributions that are known only up to a constant of proportionality. Suppressing the data and other parameters from the notation, we have, by the definition of conditional probability,

$$\mathrm{P}\left(\breve{\psi}_k | \breve{\boldsymbol{\psi}}^{-k}\right) \propto \mathrm{P}(\breve{\boldsymbol{\psi}}).$$

Thus, the joint posterior (or a function proportional to it) can be used as the unnormalized density of the conditional posterior distributions of each element of $\breve{\boldsymbol{\psi}}$ conditional on the others.

The complete MCMC routine is:

1. Choose initial values $\breve{\boldsymbol{\psi}}$ for the parameters of the underlying TBVN distribution of $\beta^b$ and $\beta^w$.
2. Draw values from the posterior distributions of $\boldsymbol{\beta}^b$ conditional on the current values of $\breve{\boldsymbol{\psi}}$, and the data, using inverse CDF sampling from these TN distributions.
3. Draw new values for each element of $\breve{\boldsymbol{\psi}}$ conditional on the previous values of the others, $\boldsymbol{\beta}^b$, and the data, using ARMS.
4. Repeat from step 3.

### 4.2.2 Ecological Inference in Several Elections at Once

I now extend King's EI and the MCMC procedure to the case in which multiple elections are observed for the same set of geographic units (precincts).[10] In this extended model, precinct-level estimates of $\beta^b$ and $\beta^w$ for each of a series of elections are improved through the borrowing of strength, not only across precincts within elections, but also across elections within the same precinct.

[9] Computer routines implementing ARMS from user-written density functions are provided by Gilks et al. at `http://www.mrc-bsu.cam.ac.uk/pub/methodology/adaptive_rejection/`.
[10] Computer programs for estimating King's basic EI model using MCMC are available from the author.

Consider as set of elections $j = 1, 2, \ldots, J$ held in set of precincts $i = 1, 2, \ldots, p$. All of the general features of the EI model described above are maintained. In particular, the joint distribution of $\beta^b_{ij}$ and $\beta^w_{ij}$ is assumed to be bivariate truncated normal and independent of the $X_{ij}$. The identity

$$T_{ij} = X_{ij}\beta^b_{ij} + (1 - X_{ij})\beta^w_{ij}$$

holds. The parameters describing the joint distribution of $\beta^b_{ij}$ and $\beta^w_{ij}$ ($\breve{\psi}_{ij}$) are the following:

$$\begin{aligned}
\breve{\mathfrak{B}}^b_{ij} &= \bar{\mathfrak{B}}^b_j + \mu^b_i,\\
\breve{\mathfrak{B}}^b_{ij} &= \bar{\mathfrak{B}}^w_j + \mu^w_i,\\
\breve{\sigma}^b_{ij} &= \breve{\sigma}^b_j,\\
\breve{\sigma}^w_{ij} &= \breve{\sigma}^w_j,\\
\breve{\rho}_{ij} &= \breve{\rho}_j.
\end{aligned}$$

The location of the TBVN distribution is a function of fixed precinct-specific and election-specific components. The dispersion parameters and correlation parameter have only election-specific components.[11] In order to separately identify the precinct and election location parameters, the expectations of the precinct location effects are assumed to be 0. In particular, I assume

$$\left(\mu^b_i, \mu^w_i\right) \sim \mathrm{BVN}(\mathbf{0}, \mathbf{\Sigma})$$

for $i = 2, 3, \ldots, p$, where[12]

$$\mathbf{\Sigma} = \begin{bmatrix} \omega^2_b & 0 \\ 0 & \omega^2_w \end{bmatrix}.$$

The hyperparameters describing the variances of the precinct effects, $\omega^2_b$ and $\omega^2_w$, are given inverse chi-square priors.

The basic MCMC procedure described above is maintained, except that additional steps to allow Gibbs sampling from the conditional distributions of the additional parameters are added. The expanded procedure is:

1. Choose initial values $\breve{\psi}_j$ for $j = 1, 2, \ldots, J$ for the parameters of the underlying TBVN distribution of $\beta^b$ and $\beta^w$.
2. Choose initial values for $\mu^b_i$ and $\mu^w_i$ for $i = 1, 2, \ldots, p$.
3. Draw values from the posterior distribution of $\boldsymbol{\beta}^b_j$ conditional on $\breve{\psi}_j$ for $j = 1, 2, \ldots, J$, $\mu^b_i$ and $\mu^w_i$ for $i = 1, 2, \ldots, p$, and the data, using inverse CDF sampling from these TN distributions.
4. Draw new values for each element of $\breve{\psi}_j$, for $j = 1, 2, \ldots, J$ conditional on the previous values of the others, the current values of $\boldsymbol{\beta}^b$, $\mu^b_i$ and $\mu^w_i$ for $i = 1, 2, \ldots, p$, and the data, using ARMS.

[11] Precinct-specific dispersion and correlation parameters are feasible, though using them adds substantial computational burden. Because the number of elections is typically small, the posterior distributions of the precinct variances and correlation components are unlikely to be very informative. If, on the other hand, one observed many elections in a small number of precincts the $i$ and $j$, subscripts might reasonably be interchanged.

[12] The precinct effects are assumed to be independently drawn across precincts.

5. Draw new values for each element of $\mu^b$ and $\mu^w$ conditional on the other parameters and the current values of $\beta^b_j$ for $j = 1, 2, \ldots, J$ and the data, using ARMS.
6. Draw new values for $\omega_b$ and $\omega_w$ conditional $\mu^b_i$ and $\mu^w_i$ for $i = 1, 2, \ldots, p$ from the appropriate inverse chi-square distribution.
7. Repeat from step 3.

As noted by King (1997), there is relatively little information in the data about the parameters $\rho_j$ for $j = 1, 2, \ldots, J$ that describe the correlation between $\beta^b_{ij}$ and $\beta^w_{ij}$. In what follows, I restrict $\rho = 0$. The assumption that $\beta^b_{ij}$ and $\beta^w_{ij}$ are a priori independent is widely assumed in the literature (see, for example, King, Tanner, and Rosen, 1999, Introduction, p. 8; or Wakefield, 2001). This restriction greatly reduces the computational burden and numerical problems associated with the estimation.

Because the posterior distribution of the elements of $\breve{\psi}$ are highly correlated (particularly if the degree of truncation is large), the MCMC routine converges slowly. Additionally, numerical failure of the bivariate normal density call can occur if the degree of truncation becomes too large.[13] To avoid these problems the values of $\breve{\mathfrak{B}}^b_j$ and $\breve{\mathfrak{B}}^w_j$ are assumed to lie in the interval $[-0.5, 1.5]$. This restriction is applied through a uniform prior on $[-0.5, 1.5]$ interval for these parameters. Such a large restriction on the possible values of these theoretically unbounded parameters needs to be justified. In the next section, I demonstrate that truncated normal distributions with location parameters outside $[-0.5, 1.5]$ can be very closely approximated by truncated normal distributions with location parameters in that interval.

### 4.3 ESTIMATING THE TRUNCATED BIVARIATE NORMAL PARAMETERS WHEN THE DEGREE OF TRUNCATION IS LARGE

One of the main technical difficulties in implementing King's EI revolves around the estimation of the parameters of the truncated bivariate normal distribution when one or both of the location parameters are not in the interval (0, 1). Figure 4.1 illustrates this problem in the simple case where the $\rho = 0$ and thus $\beta^b$ and $\beta^w$ follow univariate truncated normal distributions. The solid lines in the figure show the density over the unit interval when the location ($\mu$) of the TN distribution is 2 and the dispersion ($\sigma$) is 0.25, 0.5, and 1.0. The dashed lines show the most similar TN distributions with location parameters equal to 5. The dotted lines show the most similar TN distributions with location parameters equal to 1.25. Note that in each case, the solid line is closely approximated by the dashed and dotted lines despite the disparity in the location parameters of the underlying distributions. Even if a large number of direct observations on $\beta^b$ were available, it would be very difficult to infer the exact location and spread parameters of the underlying distribution. In the EI model, $\beta^b$ is not directly observed. Uncovering the differences in the densities shown in Figure 4.1 through EI involves detecting small differences in *latent* distributions.

In and of itself, the fact that the likelihood will be locally very flat and skewed away from the unit interval when the true location parameter is not in the unit interval does not present a problem. However, in this case calculations of the likelihood becomes increasingly inaccurate as the estimated location parameter is moved off the unit interval. Thus, both the maximization procedures used by King and the MCMC techniques presented here can become unstable if the location parameters are allowed to stray too far from the unit interval.

[13] When $\rho = 0$ is imposed, the bivariate normal call becomes the product of univariate cumulative normal calls, greatly reducing the numerical inaccuracies.

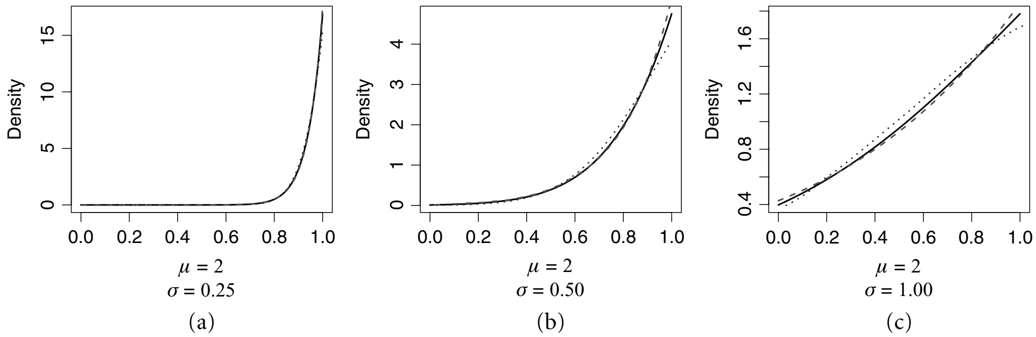



**Figure 4.1.** Discerning between truncated normal distributions. Truncated normal distributions with location parameters that lie beyond the support of the distribution can be closely approximated by other *TN* distributions. The solid line in each figure shows a *TN* density with location parameter equal to 2. The dashed line shows the closest *TN* density with location parameter equal to 5, and the dotted line shows the closest *TN* density with location parameter equal to 1.25.

On the other hand, the fact that TN distributions with very different parameterizations yield very similar densities implies not only that these parameters are difficult to estimate, but also that their exact values are not required to calculate the ultimate quantities of interest. These quantities of interest, such as the fraction of blacks that vote in each precinct or district-wide, are determined by densities that can be accurately estimated even if the parameters of the TN distribution cannot.

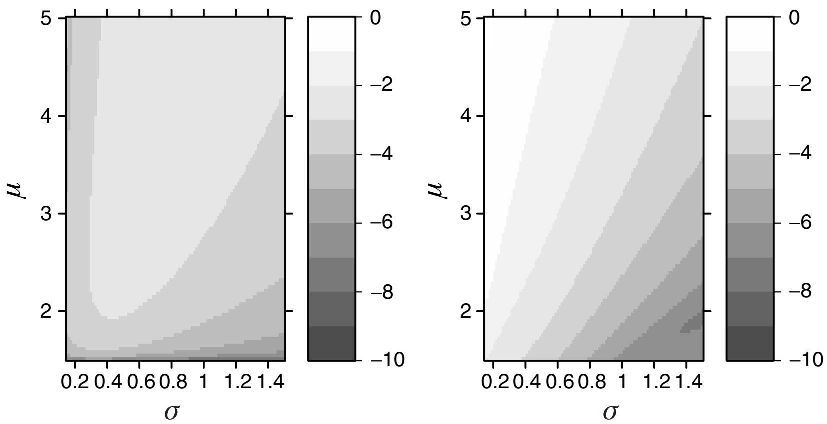
In what follows the values of the location parameters are restricted to fall between −0.5 and 1.5. This effectively avoids the numerical inaccuracies that arise when more extreme regions of the posterior density of the TBVN distribution are investigated, without appreciably affecting the posterior distributions of the precinct-level parameters of interest. Figure 4.2 shows how closely TN distributions with location parameters at 1.5 can approximate TN normal distributions with various location and dispersion parameters. The distance between distributions is measured by the Kullback–Liebler distances (Kullback and Liebler, 1951). The Kullback–Liebler distance between the true density $f$ and the approximate density $g$ is

$$I(f, g) = \int \ln\left(\frac{f(x)}{g(x)}\right) f(x)\, dx.$$

The distance $I$ is commonly interpreted as the expected value of a likelihood-ratio test which attempts to discriminate between $f$ and $g$ using a single observation. The distances shown in Figure 4.2 are typically about $10^{-3}$, often smaller, and in no case greater than $10^{-2}$. By way of comparison, Figure 4.2 also shows the distances between the same set of TN distributions and truncated Student's $t$ distributions with 80 degrees of freedom.[14] The truncated Student's $t$ distribution with 80 degrees of freedom is chosen as a basis of comparison because its very close similarity to the normal is well known.[15] While the quality of the truncated Student's $t$ approximation to the TN distributions is more variable, the overall quality of the

[14] Truncated Student's $t$ distributions with the same location and dispersion parameters as the corresponding TN distributions are used for these comparisons.

[15] This comparison may be somewhat misleading because the region of truncation is often in the extreme tails where the Student's $t$ and normal distribution differ most greatly. However, other similar heuristic comparisons yields similar results. For example, untruncated normal distributions with unit variance and means that differ by 0.045 have $I = 1 \times 10^{-4}$.




(a) (b)

**Figure 4.2.** Level plots of Kullback–Liebler distances between truncated normal distributions on the interval (0, 1) with the given parameters and (a) the closest truncated normal distribution with $\mu = 1.5$, (b) the corresponding Student's $t$ distributions with 80 degrees of freedom. The scales of Kullback–Liebler distances are the order of magnitude ($\log_{10}$).

approximation is similar to that found when the TN with location parameter at 1.5 is used to approximate TN distributions with larger location parameters. Similar values for $I$ are given by Aitchison and Shen (1980) for logistic normal approximations to Dirichlet distributions and are taken as evidence that logistic normal models can very closely approximate Dirichlet data.

## 4.4 APPLYING SUEI TO SIMULATED DATA

In this section, I report the results of the application of the SUEI estimator to simulated data that follow the probability models described above. These simulations reveal (1) how the gains from SUEI vary as a function of the number of observed elections, (2) the correlation in $X$ with precincts over time, and (3) when the conditions for aggregation bias are present in the data.

### 4.4.1 Validating the SUEI Model

I first consider simulated data that include five elections and 250 precincts. The main parameters of the TVBN distributions are $\mathfrak{B}^b = (1.00, 0.30, 0.60, 0.70, -0.10)$, $\mathfrak{B}^w = (0.40, 0.70, 0.40, 0.00, 0.30)$, $\boldsymbol{\sigma}^b = (0.15, 0.07, 0.10, 0.05, 0.10)$, and $\boldsymbol{\sigma}^w = (0.05, 0.07, 0.04, 0.12, 0.04)$.[16] The values of $X$ are drawn from a uniform distribution on the interval [0, 1/2] and are fixed across the elections within precincts, as would typically be the case with data on racial composition by precinct. The precinct effects are distributed normally across the precincts with a mean of 0 and standard deviations equal to 0.15 for $\mu^b$ and 0.10 for $\mu^w$.

The posterior distributions of the estimated main truncated biviariate normal distributions are shown in Figure 4.3.[17] The "true" values of these parameters are shown as dots on

[16] The $\boldsymbol{\rho}$ parameters are all set equal to zero.

[17] These plots and other results presented are based on 500,000 MCMC iterations, of which the first 100,000 are discarded.

P1: FZZ/FZZ P2: FZZ
CB658-04drv CB654-KING-Sample CB658-KING-Sample.cls January 21, 2004 0:32



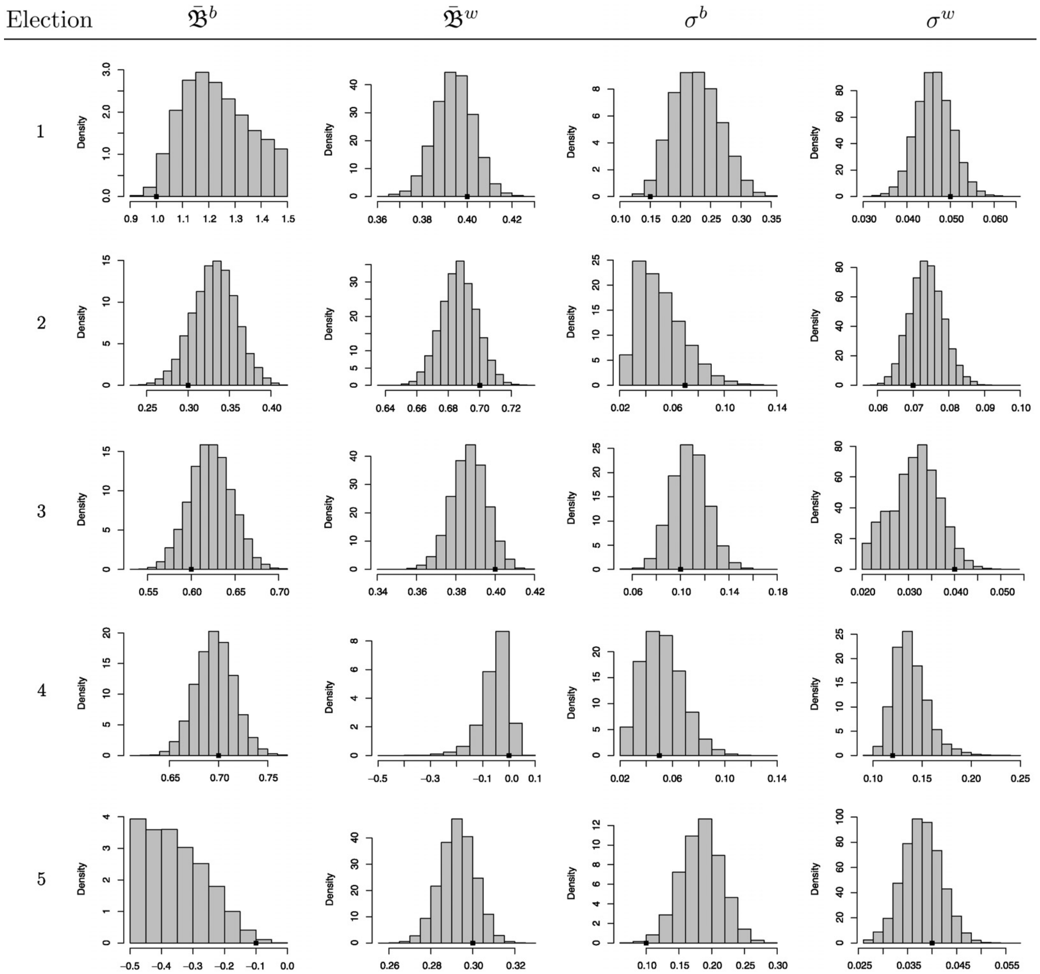
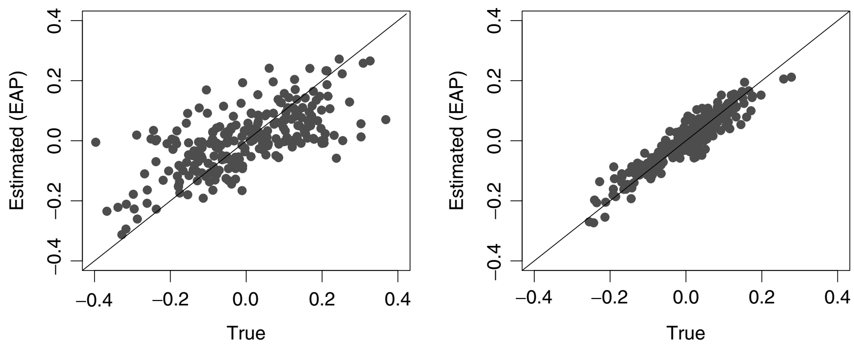



**Figure 4.3.** Estimates of model parameters from simulated data. The plots show histograms of the estimated posterior probabilities of of the model parameters. The dots on the axis of each graph indicate the "true" values of each parameter in the simulated data.

the axis of each histogram. In most cases, they fall near the bulk of the posterior mass. In few cases, they fall fairly far from the mass – in particular, in the case of $\mathfrak{B}^b$ for elections 1 and 5, where the true $\mathfrak{B}^b$'s are 1 and 0 respectively. In both cases, the posterior distributions lie mainly to the extreme side of the true value. Given the strong negative collinearity between $\mathfrak{B}^b$ and the corresponding $\sigma^b$ when $\mathfrak{B}^b$ lies off (or, in this case, on the boundary) of the unit interval, the fact that the true $\sigma^b$'s fall on the left edge of the posterior distributions associated with elections 1 and 5 comes as little surprise. However, it should be noted that for $\mathfrak{B}_4^w$, whose true value is 0, the posterior mass is much closer to true value. Here again the distribution is severely skewed away from the unit interval. As expected, the data are able to place low posterior probability on the values of $\mathfrak{B}_4^w$ that lie in the unit interval, but place relatively more weight on extreme values out of the unit interval.[18] While not shown in

[18] A key question is whether these posteriors are evidence that the MCMC estimator has not converged. However, there is little evidence that this is the case. Using a variety of starting values and rerunning the estimator




(a) (b)

**Figure 4.4.** Estimated versus actual precinct effects: plots of the posterior mean estimates of the precinct effects against their "true" values in the simulated data: (a) nonwhite, (b) white.

Figure 4.3, the MCMC estimator is very effective at recovering the variation in the precinct effects, estimating $\omega_b$ to be 0.153 with a 95 percent credible interval of (0.134, 0.172), and $\omega_w$ to be 0.095 with a credible interval of (0.088, 0.112).

The expected a posteriori (EAP) precinct effects for each precinct in the simulated data are plotted against the true values in Figure 4.4. Both the true and estimated precinct effects for the nonwhite precinct populations exceed those from the white groups, as follows from the data, in which $\omega_b = 0.15$ and $\omega_w = 0.10$. The estimates of the $\mu^w$ generally correspond more closely to the true value. This is because the white group is considerably larger than the nonwhite group in most precincts, and thus the logical bounds on the precinct fractions of white turnout are typically tighter than those for nonwhite turnout.

The more important – and indeed central – question addressed by the simulation is how much improvement in the estimation of the ultimate quantities of interest result from the incorporation of precinct effects. Table 4.1 addresses this question. Here results of the SUEI model are compared with results of using King's EI estimator on each of the five sets of election data separately. This is not a perfect comparison, because the assumptions of SUEI and King's basic EI are not nested unless there are no precinct effects ($\omega_b = \omega_w = 0$). If truncation on the unit square is negligible, the two models are nearly nested. That is, the distribution $\beta^b$ and $\beta^w$ in each precinct will be a normal mixture (determined by the unobserved and in King's EI unidentified precinct effects) of nearly normal variables (the $\beta$'s themselves). Because normal mixtures of normal variables are also normally distributed, without truncation King's EI and SUEI will be nested and the distributional assumptions of both models will simultaneously be satisfied. With truncation, however, this is no longer the case. The normal mixture of truncated normals that is the assumed distribution of $\beta^b$ and $\beta^w$ in SUEI is not the truncated normal distribution required for King's EI.[19] However, if the degree of truncation is relatively small or if the variance in the mixture that arises from the precinct effects is small relative to the election specific variation, the degree to which data generated under the SUEI assumptions differ from data generated under the standard EI assumptions will be relatively small. In these data, deviations of the simulated

consistently yielded similar posteriors. Applying King's estimator to the simulated data for a single election sometimes yields point estimates closer to the true modes; however, using the MCMC estimator on one election produces results similar to King's EI, suggesting that the difference results from the introduction of the precinct effects and not the MCMC procedure itself.

[19] This is because normal mixtures of truncated normals are not truncated normals.

**Table 4.1** Estimated quantities of interest for the simulated data

| Election | | District-wide | | Precinct-level std. dev./MSE | |
|---|---|---|---|---|---|
| | | $B^b$ | $B^w$ | $\beta^b$ | $\beta^w$ |
| 1 | Truth | 0.86 | 0.40 | 0.11 | 0.10 |
| | Basic EI | 0.86 | 0.40 | 0.099 | 0.043 |
| | SUEI | 0.87 | 0.40 | 0.088 | 0.033 |
| 2 | Truth | 0.31 | 0.70 | 0.16 | 0.12 |
| | Basic EI | 0.33 | 0.69 | 0.148 | 0.061 |
| | SUEI | 0.34 | 0.69 | 0.127 | 0.050 |
| 3 | Truth | 0.60 | 0.40 | 0.18 | 0.10 |
| | Basic EI | 0.62 | 0.39 | 0.150 | 0.059 |
| | SUEI | 0.62 | 0.39 | 0.122 | 0.043 |
| 4 | Truth | 0.70 | 0.10 | 0.15 | 0.09 |
| | Basic EI | 0.70 | 0.10 | 0.120 | 0.043 |
| | SUEI | 0.69 | 0.10 | 0.111 | 0.037 |
| 5 | Truth | 0.07 | 0.30 | 0.07 | 0.10 |
| | Basic EI | 0.06 | 0.30 | 0.071 | 0.030 |
| | SUEI | 0.08 | 0.29 | 0.070 | 0.030 |

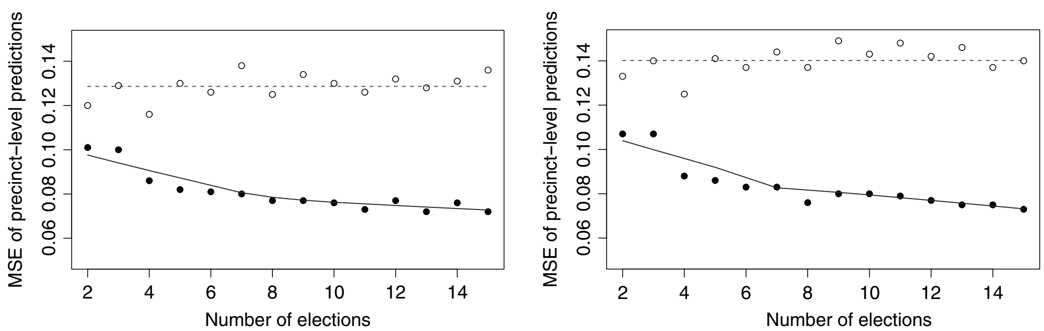
*Note:* "Truth" rows give actual district-wide quantities and the actual standard deviation of the precinct-level quantities. The other rows give expected a posteriori estimates or mean square errors of those estimates across precincts.

data from the TBVN are sufficiently small that any observed differences between the EI and SUEI estimates do not follow from the fact that the simulated data were generated in a way that is not strictly consistent with the assumptions of King's EI.

Table 4.1 reveals that MSEs of the EAP estimates of the precinct quantities of interest are consistently smaller for SUEI than for King's EI. That is, as one would expect, borrowing strength improves the predictions of the precinct quantities. The gains are, however, modest. MSEs for $\beta^w$ point estimates from the SUEI model are on average 11 percent smaller than the basic EI estimates; they are never larger, and at best are 19 percent smaller. For $\beta^w$ the percentage improvements in the MSE of SUEI over basic EI are somewhat larger than for $\beta^w$, averaging 16 percent smaller, never larger, and at best 27 percent smaller. While these these improvements are not huge, they are nonnegligible.

### 4.4.2 Investigating SUEI Efficiency Gains

I investigated how SUEI performed versus EI in three simulated data experiments. In the first experiment, I varied the number of observed elections. In the second experiment, I varied the correlation in $X$ within precincts across elections. In the third experiment, I investigated the robustness of SUEI to aggregation bias. In all of these experiments, the same set of TBVN parameters was used for every election; $\breve{\psi} = (0.8, 0.4, 0.1, 0.1, 0.0)$. The percent white $(X)$ is assumed to be uniformly distributed over the interval $(0, 1)$ across precincts in each election. The number of precincts $p$, is set to 150. The standard deviation




(a) (b)

**Figure 4.5.** The average precinct-level mean squared error of (a) the $\beta^b$ estimate and (b) the $\beta^w$ estimate, as a function of the number of observed elections across 14 simulated data sets as described in the text.

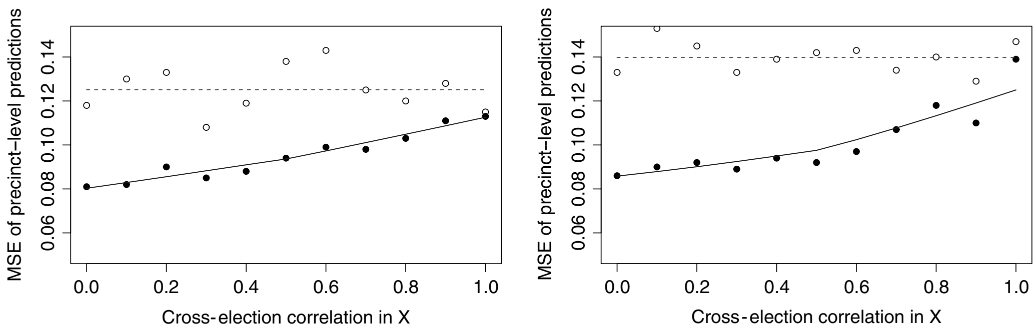
of the precinct effects was set to 0.2 for both the white and nonwhite groups in each case. Large gains from SUEI would in some cases be possible if the TBVN parameters varied from election to election, but if we use the same probability model to generate each election in a given experimental trial, the results are easy to compare across methods and experiments.[20] In particular, because the marginal distribution of $X$ and the parameters of the TBVN are the same in every election considered in all of these three experiments, the EI estimates should only vary as a function of sampling. On the other hand, as the number of elections is increased or as the correlation in $X$ across elections decreases, the efficiency of SUEI should increase.

I begin by constructing a series of simulated data sets with the given parameter values. The first data set contains two elections, the second three elections, and so forth, up to the largest data set, which contains 15 elections. In contrast to the simulated data set presented in the previous subsection, here the values of $X$ are independently drawn across elections, which (as shown below) increases the efficiency of SUEI estimates relative to the case in which $X$ is fixed across elections.

Figure 4.5 shows the average MSE of the precinct-level quantities of interest across all precincts and elections for each of the data sets. The open circles show the MSE for standard EI estimates; the solid circles, for the SUEI estimates. The dashed line plots the trend in average MSE of the EI precinct-level estimates as the number of elections in the data is varied. The solid line plots the trend in the MSE of the SUEI estimates as the number elections is varied. Notice the dashed line is flat, reflecting the fact that EI does not borrow strength across elections. However, the quality of the SUEI estimates increases as more elections are observed and more information is pooled. Even when only two elections are observed, SUEI yields MSEs that are about 15 to 20 percent smaller than those produced by EI. With 15 observed elections the reduction in MSE approaches 50 percent. The graphs reveal diminishing returns to each additional observed election. Given the variances of the election-specific and precinct-specific components and leaving aside the truncations, the upper bound of the reduction in the precinct level MSE is approximately 55 percent.[21]

[20] The SUEI estimates are based on 100,000 iterations of the MCMC routine (the first 5000 iterations are discarded).

[21] Leaving aside truncation, the MSE of the precinct quantities in EI would be $\sqrt{0.2^2 + 0.1^2} = 0.23$, as opposed to 0.10 if the precinct effects were known. Due to truncation, the MSEs are lower (about 0.13 for EI and 0.07 for SUEI with 15 elections).




(a) (b)

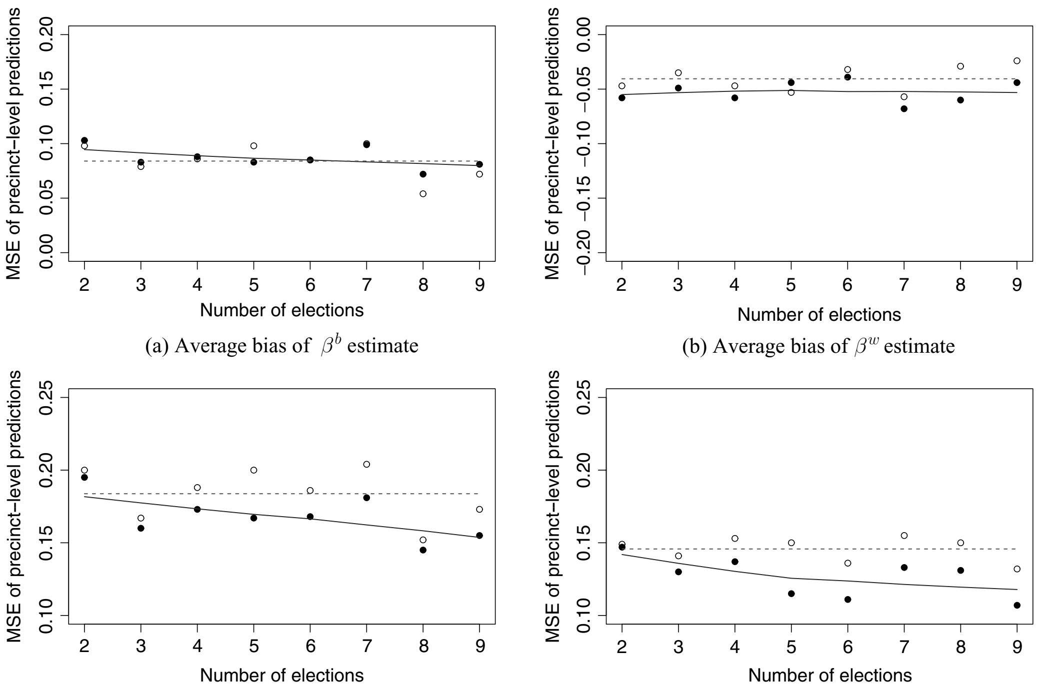
**Figure 4.6.** The average precinct-level mean squared error of (a) the $\beta^b$ estimate and (b) the $\beta^w$ estimate as functions of the correlation in $X$ across elections as described in the text.

In the previous experiment, $X$ was drawn independently across elections. Allowing $X$ to vary within precincts across elections increases our ability to infer the values of the precinct-level effects and in part the explains why I find larger gains in efficiency using SUEI in the this experiment than I did in the first simulated data set presented in the previous section. To see how the advantage of SUEI over EI varies as a function of the variation in $X$ across elections, I created 11 simulated data sets. Each of the data sets includes eight elections generated by the same parameter values as the previous experiment, with one exception. In each of the data sets, the $150 \times 8$ matrix $\mathbf{X}$ is constructed by drawing from the distribution

$$\mathbf{X}_i^* \sim \text{MVN}(\mathbf{0}, \mathbf{S}) \qquad \text{for } i = 1, \ldots, p$$

for each precinct independently, where the $8 \times 8$ variance matrix $\mathbf{S}$ has ones along the main diagonal and $r \in [0, 1]$ in each of the off-diagonal entries. Thus, the pairwise correlation between any two columns of $\mathbf{X}^*$ is $r$. I then create $\mathbf{X}$ by taking the inverse standard normal CDF of each element of $\mathbf{X}^*$. Across the 11 data sets, $r$ is varied from zero ($X$ is drawn independently across elections) to one ($X$ is constant across elections). By construction, the marginal distribution of $X$ in every election across the 11 data sets is uniform on the interval $[0, 1]$ . Thus, as in the previous experiment, the (marginal) probability model generating each election is identical, and EI, which does not pool information across elections, should generate similar estimates for each election, regardless of the correlation in $X$ across elections.

Figure 4.6 plots the MSEs of the estimates of precinct quantities of interest across the 11 simulated data sets. The dotted line representing the trend in the EI MSEs remains flat as the correlation in $X$ across elections is increased. The advantage of SUEI over EI is greatest when $X$ is drawn independently across elections, and least when $X$ is identical across elections. This result follows from the fact that precinct effects can be more precisely estimated when there is variation in $X$ across elections. Without variation in $X$ (and without variation in the main parameters of TBVN across elections), SUEI can still recover some information about the precinct effects, in cases in which $T$ is consistently higher or lower than average across elections; without variation in $X$, however, there is little information in the data to separate the overall pattern in turnout into nonwhite ($\mu^b$) and white ($\mu^w$) components. Nevertheless, the experiment reveals efficiency gains of 5 percent even when $X$ and the parameters of the TBVN are constant across elections (the least favorable conditions for borrowing strength across elections).




(a) Average bias of $\beta^b$ estimate (b) Average bias of $\beta^w$ estimate

(c) Average MSE of $\beta^b$ estimate (d) Average MSE of $\beta^w$ estimate

**Figure 4.7.** Average bias and MSE of precinct-level estimates as a function of the number of observed elections.

In a last set of experiments, I consider whether SUEI is more robust to data that violate the independence assumption, which is critical to avoiding bias in EI or ecological regression. In these experiments, I created data sets containing between two and nine elections which followed the same probability model as the previous experiments except that the cross-election correlation in $X$ was fixed at 0.7, and $\breve{\mathfrak{B}}^b_{ij} = 0.8 + 0.4(X_{ij} - 0.5)$. Figure 4.7 shows the average bias and MSE of the EI and SUEI as a function of the number of elections. The top two panels reveal that SUEI was no more robust to aggregation bias than EI. When $\breve{\mathfrak{B}}^b$ (and, thus, $\beta^b$) is a function of $X$, estimates of $\beta^b$ and $\beta^w$ are biased. Increasing the number of elections does not reduce the bias in the SUEI estimates. However, the lower two panels reveal that even in the presence of bias, SUEI still reduces the MSE of the precinct-level prediction versus EI, and that advantage increases with the number of observed elections.

In other experiments, I ran SUEI and EI on data sets which included some elections in which the conditions for aggregation were present as well as some in which those conditions were not present. In those experiments, SUEI did somewhat decrease in the bias of the precinct-level estimates relative to EI, though the differences were not dramatic. The larger advantage of SUEI over EI when the independence assumption is violated may be found if the SUEI model is extended to allow the $\breve{\mathfrak{B}}^b$ and $\breve{\mathfrak{B}}^w$ to depend on $X$ as in the extended EI model. I leave this extension to be investigated in future work.

While the results of these simulations are not definitive, they do yield some important observations. As the number of elections considered increases, the advantage (in terms of MSE) of SUEI over EI grows (to as much as 45 to 50 percent). Similarly as the racial compositions of the districts becomes more variable across elections, the advantage of SUEI grows, although some advantage is found even if $X$ is fixed across elections. On the the other hand, I would have found smaller reductions in MSE from SUEI if the estimated precinct effects had been smaller relative to the election-specific effects.[22]

## 4.5 TURNOUT BY RACE IN VIRGINIA ELECTIONS

Virginia presents a good example of a setting in which ecological inference might be improved through consideration of several elections at once. Because Virginia elects its governors to four-year terms in odd-numbered years and its entire State senate in the odd-numbered years which do not have gubernatorial elections, whereas federal elections are in even-numbered years, important state or federal contests are held in Virginia every year. Thus, in a short period of years – over which geographic composition might safely be assumed to be stable – a sizable number of significant elections are held. I consider an example drawing on data from the 1984 through 1990 Virginia elections.[23] The object of inference is the rate of turnout among whites and nonwhites, which will be estimated for each of the seven elections. It is not possible to obtain direct measures of turnout by racial group.[24] However, there exist previous estimates and expert opinions which can be used as points of comparison.

The question of turnout by race in Virginia elections in the 1980s is of particular interest (see, for example, Hertzog, 1994; Sabato, 1987, 1991; Strickland and Whicker, 1992; Traugott and Price 1992; Schexnider 1990; Morris and Bradley, 1994). The 1980s saw the emergence of African-American candidates for statewide office in Virginia and the nation. In 1985 Virginians elected an African-American, L. Douglas Wilder, lieutenant governor, and in 1989 they elected him governor. In 1988, the Republicans nominated Maurice Dawkins, an African-American, for the U.S. Senate. In 1984, Jesse Jackson won the Democratic caucus vote in Virginia (though he ran second to Mondale in national conventional delegates), and in 1988, Jackson captured a plurality (45 percent) of the Democratic primary vote. Thus, Virginia in the 1980s offers an interesting testing ground for theories about the electoral significance of race and, in particular, the effect of minority candidates on minority-voter mobilization.

An established literature presents theoretical foundations and empirical tests of the assertion that the race of candidates or office holders affects the political mobilization of racial minority and majority groups. For example, Tate claims that black participation is generally higher when black candidates are on the ballot, though her survey evidence suggests that most blacks disagree with the assertion that "blacks should always vote for black candidates when they run" (1994, p. 105). Nevertheless, Tate argues that high black turnout rates are often associated with precedent-setting candidacies (such as Wilder's). Bobo and Gilliam (1990) show that black political engagement is greater in cities with black mayors. Gay (2001) shows that white voter turnout is depressed and black voter turnout (sometimes) increased in districts held by black members of Congress. Kleppner (1985) reports that historically high black voter turnout was critical to Harold Washington's mayoral victory in Chicago in 1983. Similarly, high black voter turnout in states like Virginia is seen by some as critical to the success of black candidates (Strickland and Whicker, 1992).

The existing estimates of turnout by race in these elections come from Sabato (1991) and are based on turnout in 44 selected predominantly black precincts. The rate of turnout in

[22] Similarly, larger advantages would have been found if the precinct effects had accounted for a larger share of the variability in $\beta^b$ and $\beta^w$.
[23] The data are from the ROAD data project (King et al., 1997).
[24] Indeed, Virginia does not collect information about the race of voters when they register.

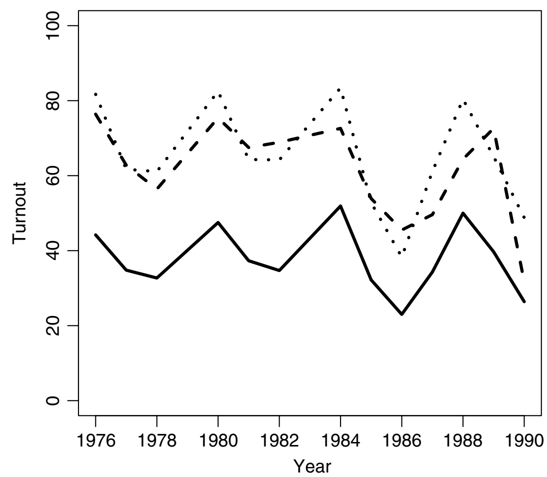



**Figure 4.8.** Estimates of black and white voter turnout from Sabato (1991). The dotted line shows white turnout, and the dashed line shows black turnout, in each case as a fraction of voter registration. The solid line shows total turnout as a fraction of the voting age population.

these precincts is taken as an estimate of black turnout statewide. Unfortunately, these estimates are for the percentage of registered voters that turn out to vote and not percentages of the total voting age populations. Because population data for the precincts are not available, the turnout rates as a fraction of voting age population cannot be estimated in a comparable way.[25] Turnout rates for blacks reported by Sabato are shown in Figure 4.8.[26] Interestingly, Sabato's results suggest that black turnout was higher than white turnout in the 1985 and 1989 races, in which Wilder was a candidate for lieutenant governor and governor. Black turnout was estimated to be lower than white turnout in 1986, when the black Republican Dawkins was a candidate for U.S. Senate, and in all of the other years in the eighties except 1981. By these estimates black turnout never exceeds white turnout by more than about 7.5 percentage points, though in some elections white turnout exceeds white turnout by as much as 17 percentage points.

In order to analyze turnout rates among whites and nonwhites using the ecological inference estimators developed above, I require election returns and racial composition data for a set of geographic units. Practically, this requires aggregating electoral returns to a level that corresponds to geographical units recognized by the Census Bureau. In the ROAD project, King et al. (1997) published electoral data for Virginia elections from 1984 to 1990 that are aggregated to the minor civil division (MCD) group level. In the main these are simply the Census Bureau's MCDs (for example, Alexandria, Berryville, or Quantico) except in cases where one or more electoral precincts (the lowest level of electoral aggregation) were shared across two or more MCDs. In these cases, the MCDs sharing precincts are grouped so that no electoral precinct is split across groupings. In total there are 257 MCD groups in the Virginia data, ranging widely in size from 506 to 183,000 voting age residents. The median Virginia MCD group has 7363 voting age residents. Nonwhites make up 22 percent of the voting age residents statewide. The distribution of the nonwhite population across the MCD groups is shown in Figure 4.9. While many of the MCD groups have very small nonwhite populations, a small number of them are majority-minority.

[25] Similarly, because registration-by-race data are not available, ecological analysis of the sort developed here cannot be undertaken on the precinct-level data.

[26] Sabato does not give turnout rates for whites. In the figure, the white turnout rate is imputed from the total turnout rate and Sabato's black turnout rate under the assumption that 18 percent of the registered voters in Virginia were black during this period.

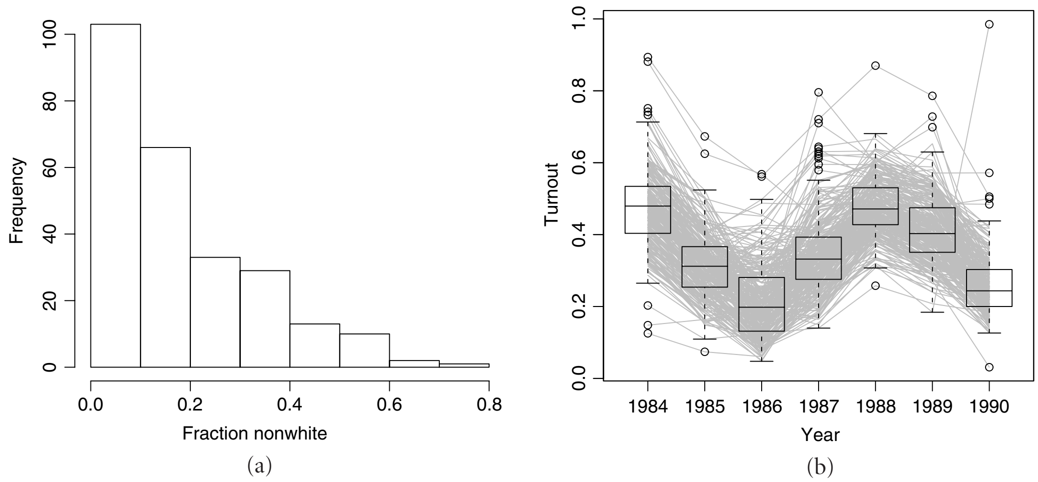



**Figure 4.9.** Virginia Ecological Election Data, 1984–1990: (a) racial composition, (b) turnout rates. Here (a) shows the distribution of nonwhite voters across Virginia minor civil division groups (MCD groups; see text for definition), and (b) shows boxplots for the turnout rates in each of the seven elections considered (as a percentage of voting age populations). Each gray line in (b) represents an MCD group.

The distribution of voter turnout across the elections is also shown in Figure 4.9.[27] The figure reveals cross-election and cross-precinct variation in overall voter turnout at the MCD-group level. As one would expect, voter turnout was highest in the presidential election years 1984 and 1988. Interestingly, the midterm elections of 1986 and 1990 had the lowest rate of turnout, even lower than the 1987 election in which no federal or statewide offices were contested. Closer inspection reveals that the 1987 election included a hotly contested statewide proposition that established the Virginia lottery, whereas the 1986 election did not involve a U.S. Senate, contest, and Senator John Warner faced no Democratic opposition in his 1990 reelection campaign (Sabato, 1991). The gray lines in Figure 4.9 trace the turnout rates within each precinct over time. Notice that there appear to be many high- and low-turnout precincts. For example, the high and low outliers tend to be the same MCD groups over time. While not sufficient to demonstrate MCD-group effects in turnout by race, persistent differences in total turnout are consistent with the existence of those effects.

Table 4.2 presents estimates of the main truncated bivariate normal parameters as estimated by King's EI and SUEI. In all but one election, the 1986 midterm, the estimated parameters are very similar. The 1984 presidential election presents a good case of what we expect to find if the data are well conditioned and the degree of truncation in the assumed TBVN distributions is small.[28] The election-specific estimated location parameters, $\mathfrak{B}^b$ and $\mathfrak{B}^w$, are identical, and the estimated election-specific standard deviations are larger for King's EI than for SUEI. This is because some of the precinct-level variation in turnout that is captured by these parameters in King's EI is attributed to the precinct effect in SUEI. Table 4.3 shows the estimated standard deviations of the precinct effects. The standard deviation of the precinct effects for both whites and nonwhites is estimated to be about 0.09. Thus, in the 1984 presidential election, the estimates are consistent with the notion that the estimated $\sigma^b$ from King's EI is decomposed to into election- and precinct-specific components

[27] Presentation of the turnout data in this way was suggested to me by James DeNardo.
[28] This should not be taken as implying that the data are in fact well conditioned. In particular, the these results are not informative about the existence of aggregation bias in the results.

**Table 4.2** Estimates of truncated bivariate normal parameters for Virginia elections data: turnout by race, 1984–1990

| | King EI | | | Precinct-effects EI | | |
|---|---|---|---|---|---|---|
| Parameter | Mean | Std. Dev. | 95% CI | Mean | Std. Dev. | 95% CI |
| *1984 Presidential* | | | | | | |
| $\breve{\mathfrak{B}}^b$ | 0.55 | 0.02 | (0.52, 0.59) | 0.55 | 0.03 | (0.49, 0.60) |
| $\breve{\mathfrak{B}}^w$ | 0.46 | 0.01 | (0.44, 0.47) | 0.46 | 0.01 | (0.47, 0.48) |
| $\breve{\sigma}^b$ | 0.12 | 0.02 | (0.08, 0.16) | 0.07 | 0.02 | (0.04, 0.10) |
| $\breve{\sigma}^w$ | 0.12 | 0.005 | (0.11, 0.13) | 0.06 | 0.004 | (0.06, 0.07) |
| *1985 Gubernatorial* | | | | | | |
| $\breve{\mathfrak{B}}^b$ | 0.47 | 0.01 | (0.45, 0.50) | 0.46 | 0.04 | (0.42, 0.51) |
| $\breve{\mathfrak{B}}^w$ | 0.27 | 0.01 | (0.27, 0.29) | 0.29 | 0.01 | (0.27, 0.30) |
| $\breve{\sigma}^b$ | 0.03 | 0.02 | (0.01, 0.07) | 0.05 | 0.01 | (0.03, 0.07) |
| $\breve{\sigma}^w$ | 0.10 | 0.003 | (0.10, 0.11) | 0.05 | 0.003 | (0.04, 0.05) |
| *1986 Midterm* | | | | | | |
| $\breve{\mathfrak{B}}^b$ | 0.25 | 0.02 | (0.21, 0.27) | −0.04 | 0.22 | (−0.44, 0.27) |
| $\breve{\mathfrak{B}}^w$ | 0.18 | 0.01 | (0.16, 0.19) | 0.19 | 0.01 | (0.18, 0.22) |
| $\breve{\sigma}^b$ | 0.09 | 0.02 | (0.05, 0.13) | 0.35 | 0.11 | (0.19, 0.53) |
| $\breve{\sigma}^w$ | 0.14 | 0.01 | (0.13, 0.16) | 0.10 | 0.01 | (0.08, 0.11) |
| *1987 State legislative* | | | | | | |
| $\breve{\mathfrak{B}}^b$ | 0.25 | 0.03 | (0.21, 0.30) | 0.20 | 0.07 | (0.09, 0.29) |
| $\breve{\mathfrak{B}}^w$ | 0.36 | 0.01 | (0.35, 0.37) | 0.37 | 0.01 | (0.35, 0.39) |
| $\breve{\sigma}^b$ | 0.10 | 0.02 | (0.04, 0.14) | 0.13 | 0.04 | (0.08, 0.20) |
| $\breve{\sigma}^w$ | 0.12 | 0.005 | (0.12, 0.13) | 0.08 | 0.01 | (0.07, 0.09) |
| *1988 Presidential* | | | | | | |
| $\breve{\mathfrak{B}}^b$ | 0.48 | 0.02 | (0.45, 0.51) | 0.48 | 0.03 | (0.43, 0.53) |
| $\breve{\mathfrak{B}}^w$ | 0.48 | 0.01 | (0.47, 0.48) | 0.48 | 0.01 | (0.47, 0.50) |
| $\breve{\sigma}^b$ | 0.10 | 0.02 | (0.07, 0.13) | 0.03 | 0.01 | (0.02, 0.06) |
| $\breve{\sigma}^w$ | 0.09 | 0.003 | (0.09, 0.10) | 0.03 | 0.003 | (0.03, 0.04) |
| *1989 Gubernatorial* | | | | | | |
| $\breve{\mathfrak{B}}^b$ | 0.55 | 0.02 | (0.52, 0.57) | 0.55 | 0.03 | (0.50, 0.60) |
| $\breve{\mathfrak{B}}^w$ | 0.39 | 0.01 | (0.38, 0.40) | 0.39 | 0.01 | (0.37, 0.40) |
| $\breve{\sigma}^b$ | 0.08 | 0.02 | (0.05, 0.12) | 0.06 | 0.02 | (0.04, 0.09) |
| $\breve{\sigma}^w$ | 0.10 | 0.003 | (0.10, 0.11) | 0.04 | 0.003 | (0.03, 0.04) |
| *1990 Midterm* | | | | | | |
| $\breve{\mathfrak{B}}^b$ | −0.13 | 0.24 | (−0.56, 0.16) | −0.10 | 0.23 | (−0.46, 0.25) |
| $\breve{\mathfrak{B}}^w$ | 0.27 | 0.01 | (0.26, 0.28) | 0.27 | 0.02 | (0.27, 0.29) |
| $\breve{\sigma}^b$ | 0.29 | 0.07 | (0.19, 0.42) | 0.29 | 0.09 | (0.15, 0.43) |
| $\breve{\sigma}^w$ | 0.10 | 0.003 | (0.09, 0.11) | 0.11 | 0.01 | (0.10, 0.12) |

*Note:* Posterior means, standard deviations, and credible intervals were calculated using King's computer procedures and the MCMC estimator described in the text.

**Table 4.3** Estimated standard deviations of the precinct-specific effects on turnout by race across the seven elections, Virginia, 1984–1990

| Parameter | Mean | Std. Dev. | 95% CI |
|---|---|---|---|
| $\omega_b$ | 0.09 | 0.04 | (0.01, 0.14) |
| $\omega_w$ | 0.09 | 0.01 | (0.08, 0.10) |

in SUEI. For example, the total nonwhite precinct-level variance is estimated in King's EI to be 0.12, and by SUEI to be $\sqrt{0.07^2 + 0.09^2} \approx 0.11$. As mentioned above, when the degree of truncation is negligible, both King's EI and SUEI imply that the precinct parameters follow bivariate normal distributions (both conditional and unconditional on the precinct effect). In such cases, precinct-level variance in King's EI will be decomposed into election- and precinct specific components as it is in the 1984 presidential election. Similar, results are obtained for the 1988 presidential election and the 1989 gubernatorial election.

In the remaining elections, differences in the estimated election-specific variance components between the two models cannot be directly attributed to the sort of decomposition described above. In these elections, the estimated election-specific variance components are larger in SUEI than in King's EI for at least one of the two racial groups. In the 1985 gubernatorial election, the EI estimated election-specific variance of $\beta^b$ is not even larger than the precinct-specific variation found using SUEI. In most cases, the differences can be attributed to greater degrees of truncation combined with differences in the ways the two models respond to violations in the their distributional assumption.

Despite differences in the estimated parameters of the underlying TBVN distributions, estimates of the aggregate quantities of interest are quite similar, as seen in Table 4.4. The maximum difference between the EI estimates and SUEI estimates are 5 percentage points for nonwhites and 1 percentage point for whites.[29] Interestingly, despite the additional efficiency that should be obtained from SUEI, the estimated posterior uncertainties in the EI estimates is generally smaller than those found for SUEI. This finding results in part from an understatement of posterior uncertainty from King's use of importance resampling and normal theory to construct estimates of the posterior uncertainty. The larger posterior uncertainties in SUEI also result from differing reactions of the two models to violations of their distributional assumptions.

The results presented in Table 4.4 support the notion that black turnout was elevated relative to white turnout in the two elections involving Douglas Wilder. In the 1985 and 1989 elections black turnout is estimated to have exceeded white turnout by about 15 to 25 percentage points. By comparison, in the 1987 state election, white turnout was estimated to exceeded nonwhite turnout by about 5 to 15 percentage points. In the two midterm elections, black and white turnout is estimated to have been quite similar. Although black turnout is estimated to have exceeded white turnout in 1986 and white turnout to have exceeded black turnout in 1990, in neither case is the difference within the 95 percent credible interval. The most anomalous case is the 1984 presidential election, in which black turnout is estimated to have exceeded white turnout by about 15 to 25 percent. While Jesse Jackson ran a strong campaign in the 1984 presidential primary, winning the Virginia caucus vote, it is not obvious that the effect of his campaign would extend to the general election six months later.

[29] That the maximum difference between EI and SUEI for whites is about 5 times smaller than for nonwhites follows directly from the fact that nonwhites comprise about 1/5 of the population.

**Table 4.4** Estimates of the statewide quantities of interest: fractions of whites and nonwhites voting statewide

| | King EI | | | Precinct-effects EI | | |
|---|---|---|---|---|---|---|
| Parameter | Mean | Std. Dev. | 95% CI | Mean | Std. Dev. | 95% CI |
| *1984 Presidential* | | | | | | |
| $B^b$ | 0.54 | 0.03 | (0.50, 0.58) | 0.53 | 0.04 | (0.46, 0.59) |
| $B^w$ | 0.41 | 0.01 | (0.40, 0.42) | 0.42 | 0.01 | (0.39, 0.44) |
| *1985 Gubernatorial* | | | | | | |
| $B^b$ | 0.47 | 0.02 | (0.44, 0.50) | 0.44 | 0.04 | (0.38, 0.50) |
| $B^w$ | 0.23 | 0.004 | (0.22, 0.24) | 0.24 | 0.01 | (0.22, 0.26) |
| *1986 Midterm* | | | | | | |
| $B^b$ | 0.25 | 0.02 | (0.21, 0.27) | 0.30 | 0.05 | (0.22, 0.39) |
| $B^w$ | 0.21 | 0.01 | (0.20, 0.22) | 0.20 | 0.01 | (0.17, 0.22) |
| *1987 State legislative* | | | | | | |
| $B^b$ | 0.25 | 0.03 | (0.20, 0.30) | 0.22 | 0.04 | (0.16, 0.28) |
| $B^w$ | 0.31 | 0.01 | (0.29, 0.32) | 0.32 | 0.01 | (0.29, 0.33) |
| *1988 Presidential* | | | | | | |
| $B^b$ | 0.46 | 0.02 | (0.42, 0.43) | 0.46 | 0.04 | (0.40, 0.52) |
| $B^w$ | 0.45 | 0.01 | (0.44, 0.46) | 0.45 | 0.01 | (0.43, 0.47) |
| *1989 Gubernatorial* | | | | | | |
| $B^b$ | 0.53 | 0.02 | (0.49, 0.56) | 0.52 | 0.04 | (0.46, 0.58) |
| $B^w$ | 0.34 | 0.01 | (0.32, 0.36) | 0.35 | 0.01 | (0.33, 0.36) |
| *1990 Midterm* | | | | | | |
| $B^b$ | 0.17 | 0.03 | (0.14, 0.22) | 0.22 | 0.04 | (0.15, 0.30) |
| $B^w$ | 0.27 | 0.01 | (0.25, 0.28) | 0.26 | 0.01 | (0.23, 0.28) |

Overall, these estimates suggest that black voter turnout is systematically higher relative to white voter turnout than Sabato's estimates suggest. Several factors might account for these differences. The 44 predominantly black precincts use by Sabato could be atypical of turnout patterns statewide. Also, Sabato assumes that nonwhite and white behavior in these precincts is the same.[30] On the other hand, it is also quite possible that there is a relationship between voter turnout and racial composition. Key's (1949) racial threat hypothesis asserts that whites will be most motivated to vote against blacks in areas where blacks are most prevalent. Consistent with Key's hypothesis, Hertzog (1994) argues that "the single most significant factor in determining how white Virginians would vote in the 1980s was the percentage of black people living the voter's locality" (p. 163). If this is true, it is quite possible that for elections in which blacks are particularly mobilized, whites in predominantly black areas will be mobilized to vote as well (for the opposing candidate). In that case, the ecological inference models considered here, which assume that racial composition and turnout by each racial group are independent, will fail in such a way that the additional white turnout in areas with large black populations will be attributed to black voters. This effect is opposite to the usual aggregation bias result, in which voting rates in predominantly black areas are lower for both blacks and whites than in predominantly

[30] Without knowing the racial composition of these precincts, the influence of white turnout on Sabato's estimates cannot be assessed.

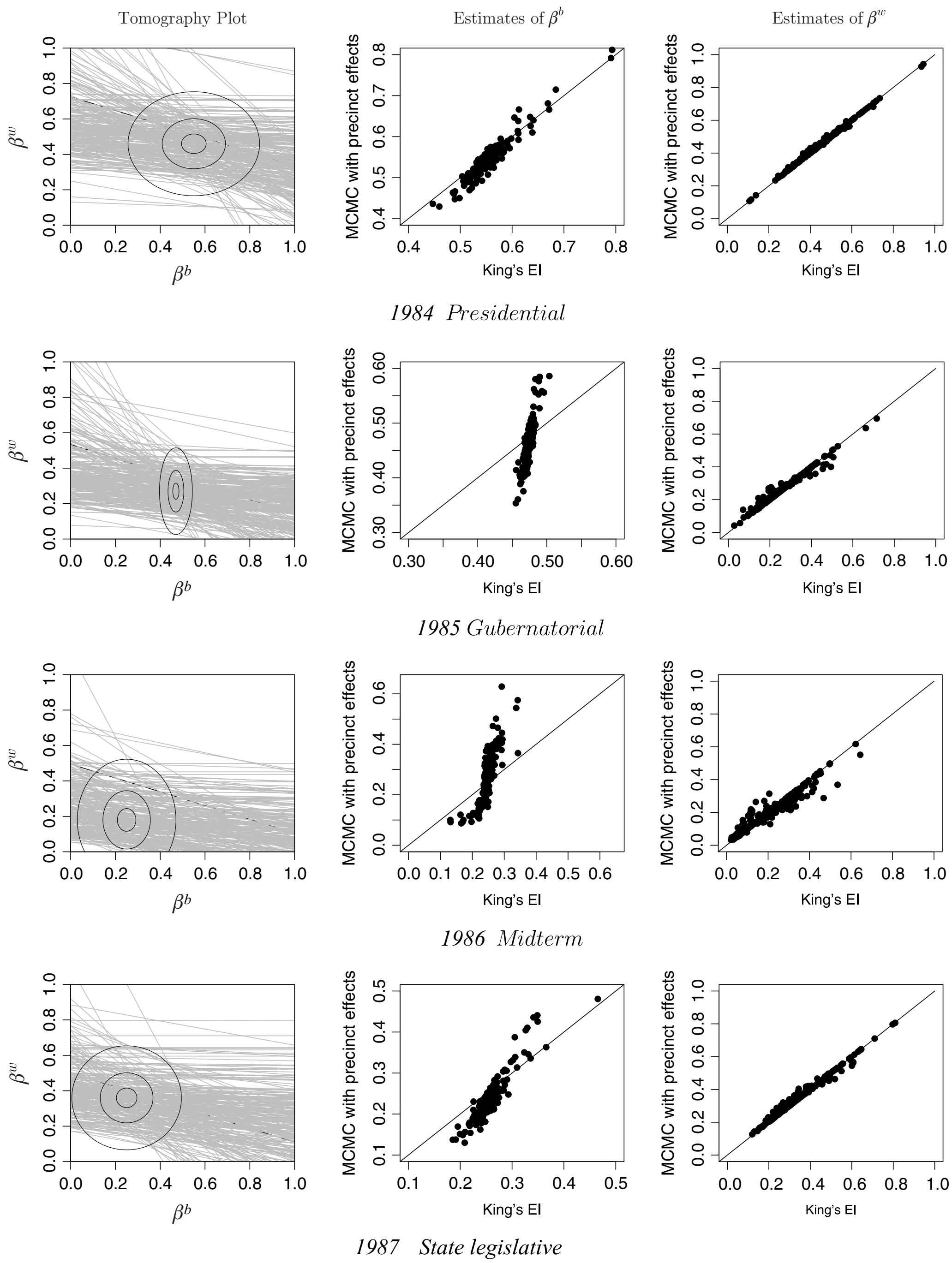



**Figure 4.10.** *MCD* group level tomography plots and estimates from King's *EI* and the precinct-effects *EI* model estimated by *MCMC*. The left panels show tomography plots of feasible values of $\beta^b$ and $\beta^w$ for each *MCD* group. The ellipses show probability contours of the *TBVN* parameters estimated by King's *EI*. The center and right panels show the *EAP* estimates of $\beta^b$ and $\beta^w$ respectively for each *MCD* group.

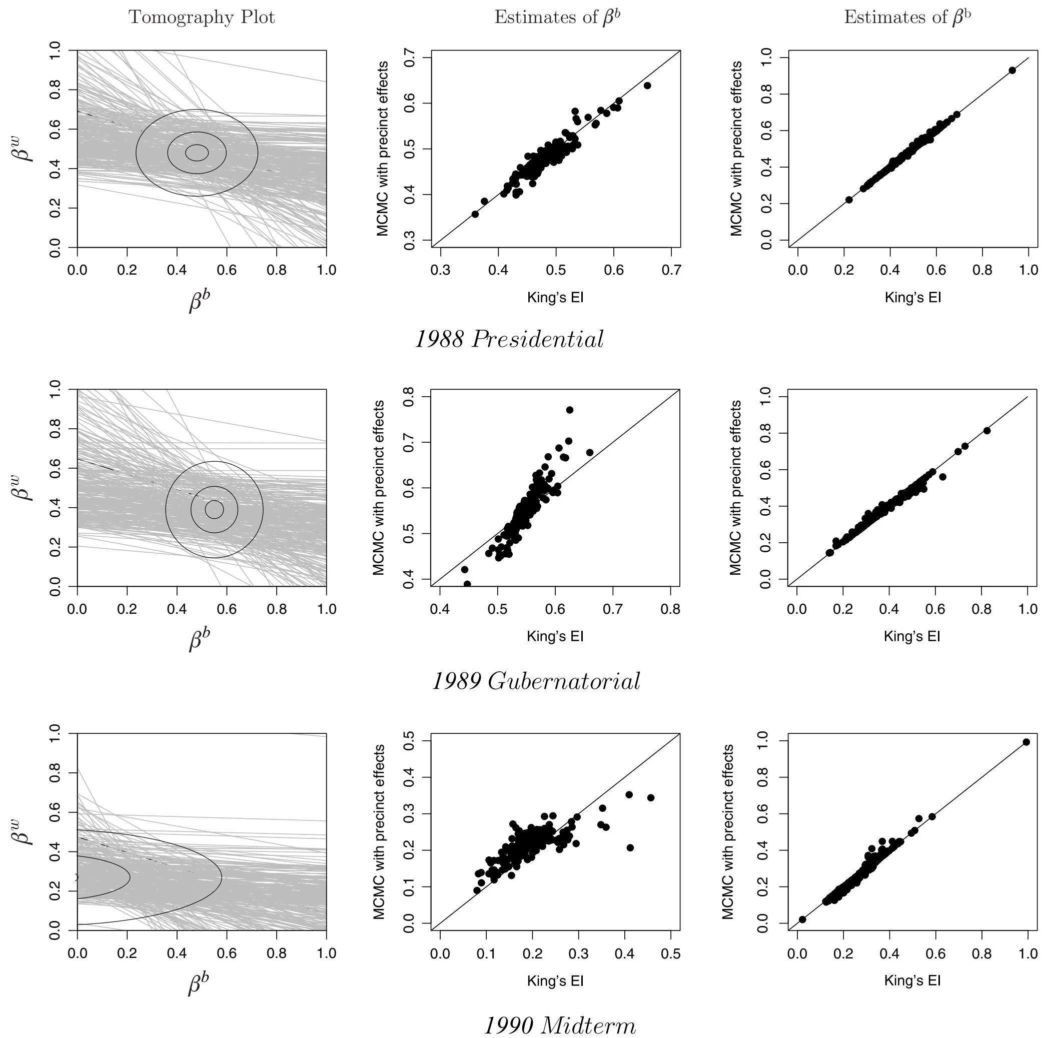



**Figure 4.10.** (*continued*)

white areas, leading the estimated black turnout to be lower relative to white turnout than is the true black turnout. Interestingly, Sabato's estimates indicates that black turnout is lower than the SUEI or EI estimates, not only in elections that involve black candidates, but in other elections as well, which undermines the idea that the differences between the two sets of estimates are due to aggregation bias resulting from racial threat. Further, the fact that Sabato's estimates which result in lower estimates of black turnout are based on the behavior of blacks (and whites) in the *most* heavily black precincts, makes less plausible the notion that there is a positive correlation between black or white turnout rates in an area and the fraction of blacks in that area. Overall, the EI and SUEI results regarding the aggregate quantities of interest are quite similar. Further, consistent with Sabato, the EI and SUEI results show higher black turnout relative to white turnout when Wilder was on the ballot.

As seen in the simulation, the real advantage in the SUEI estimator is in the improvement to the precinct-level (MCD-group-level) predictions. Figure 4.10 shows the MCD group turnout rates for whites and nonwhites as estimated by EI and SUEI along with King's so-called *tomography plot* for each election. In the tomography plot, each line represents the feasible values of black and white turnout given the total turnout rate and the racial composition in a particular MCD group. The ellipses show contour lines of the truncated normal distributions that are assumed to govern the joint distribution of white and nonwhite turnout (as estimated by EI). Notice that many of the precinct lines are very flat, indicating the feasible range of white turnout rates (plotted on the $y$-axis) is typically small and the range of feasible black turnout rates is very large (often the entire interval [0, 1]). Thus, inferring white turnout rates is a considerably easier task than inferring black turnout rates in these data. Consequently, EI- and SUEI-estimated white turnout rates in each precinct and election are quite similar, as indicated by the fact that most of the points in the white turnout ($\beta^w$) panels fall near the 45 degree lines. In the case of white turnout, borrowing strength across elections had very little effect on the estimated quantities of interest. Not that the precinct effects are not present; rather the additional information that they yield with respect to estimating white turnout rates is small. On the other hand, in several of the elections, the inclusion of precinct effects greatly increases the variation in the estimated turnout rates among blacks. That is, the posterior estimates are greatly effected by the borrowing of strength across elections. Particularly in 1985 and 1986, and to lesser extent in 1987 and 1989, SUEI finds much greater variation in black turnout than does EI. In the 1984 and 1988 elections, variation in estimated black turnout rates made by EI and SUEI are similar, and in the 1990 election the EI estimates exhibit somewhat more variation than the SUEI estimates.

Overall, when the variation in black turnout rates is estimated to be large relative to the variation in white turnout rates (when the ellipses in the tomography plots are wide), the precinct effects add relatively less, and when the variation in black turnout rates is estimated to be small relative to the variation in white turnout rates (when the ellipses in the tomography plots are tall) the precinct effect add relatively more. Also, as noted above, when the degree of truncation is large (as in 1986 or 1990), the relationship between the EI and SUEI estimates becomes more complex due the asymmetric effect that positive and negative precinct effects have on the precinct-level prediction in cases in which election specific effect ($\mathfrak{B}^b$ or $\mathfrak{B}^w$) is estimated to lie near the boundary of or off the unit square.

Of course, without knowledge of the true turnout by whites and nonwhites in each MCD group it is not possible to ascertain the degree to which the additional variation in the SUEI estimates versus the EI estimates comports with "true" cross-MCD group variation in turnout rates. However, the estimates do suggest the existence of persistent cross-election variation in turnout rates, and those difference are reflected in the SUEI MCD group data predictions. Thus, the results presented here demonstrate how the analysis of several elections at once can be used to gain leverage on the behavior of voters within each precinct (MCD group).

## 4.6 DISCUSSION

The SUEI model maintains the central assumption found in Goodman (1959) and King (1997) of independence between the turnout rates within each racial group and the racial composition of the precincts. The violation of this assumption leads to aggregation bias (Robinson, 1950) when regression-like techniques (such as Goodman's ecological regression or King's EI) are applied. While the degree to which EI is more "robust" to violations of this assumption has been debated, it is important to note the centrality of the assumption and that its violation *will* lead to bias. King presents extensions to his model in which violations of this assumption are addressed, and those same extensions could be incorporated in the model presented here. Indeed, the MCMC estimator developed above can more easily and flexibly allow for dependences between the racial composition of the precincts and the turnout rates within each group. However, there is often little information in the data to estimate such dependences (Rivers, 2000). In this regard an extended SUEI which allows for nonindependence between the precinct quantities of interest and the racial composition of the district (as King's "extended" EI) holds some promise. If the structure of nonindependence is constant across elections, then borrowing strength across elections may help to estimate that structure. This extension remains for future work.

In both the simulated data and the empirical example, the district and state-wide estimates produced by King's basic EI model and the SUEI model are very similar. The advantage of the SUEI model is in the estimates of the precinct-level quantities of interest. In the simulated data, SUEI provides improvements in mean square error of 5 to 40 percent. Because the true precinct-level quantities are not known for the Virginia election data set, the degree of improvement cannot be directly assessed. However, the estimates suggest that a considerable amount of information about precinct-level turnout by race in any given election can be gleaned from other elections.

Substantively, the estimates support the widely held, but relatively unsubstantiated, claim that nonwhite turnout exceeded white turnout in several Virginia elections in the 1980s and particularly outpaced white turnout in the 1985 and 1989 elections, in which a African-American candidate, Douglas Wilder, was on the statewide ballot. While the results may be exaggerated by an ecological fallacy if whites in areas with large nonwhite populations turned out in disproportionately large numbers to vote against Wilder (as suggested by Key's (1949) racial threat hypothesis), the general finding appears clear.

This chapter demonstrates how King's EI model can be estimated using MCMC techniques and how cross-election precinct-level dependences can be estimated and used to improve precinct-level predictions. More generally, the MCMC approach laid out in this chapter can be applied of other extensions of King's model, including perhaps ways in which the assumption of independence between the racial composition of the district and turnout rates for each group might be relaxed. Using MCMC, the posterior distributions of these tenuously identified quantities might be more accurately assessed and reliably recovered than is possible using the asymptotic normal theory and importance resampling approach described by King (1997).

## REFERENCES

Aitchison, J. and S. M. Shen. 1980. "Logistic-Normal Distributions: Some Properties and Uses," *Biometrica*, 67: 261–272.

Bobo, Lawrence, and Franklin Gilliam. 1990. "Race, Sociopolitical Participation, and Black Empowerment." *American Political Science Review*, 84, June: 277–293.

Gamerman, Dani. 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* New York: Chapman & Hall.

Gay, Claudine. 2001. "The Effect of Black Congressional Representation on Political Participation." *The American Political Science Review*, 94, 3: 589–602.

Gilks, W. R., N. G. Best, and K. K. C. Tan. 1995. "Adaptive Rejection Metropolis Sampling," *Applied Statistics*, 44: 455–472.

Goodman, Leo. 1959. "Some Alternatives to Ecological Correlation." *American Journal of Sociology*, 64: 610–25.

Hertzog, Mark Wm. 1994. "White Flight in the Voting Booth: The Racial Composition of Localities and Partisan Voting in Virginia in 1980s." In Matthew Holden, Jr. (ed.), *The Challenge to Racial*

*Stratification*. National Political Science Review, Volume 4. New Brunswick, NJ: Transaction Publishers.

Key, V. O. 1949. *Southern Politics*. New York: Knopf.

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

King, Gary, Bradley Palmquist, Greg Adams, Micah Altman, Kenneth Benoit, Claudine Gay, Jeffrey B. Lewis, Russ Mayer, and Eric Reinhardt. 1997. "The Record of American Democracy, 1984–1990." Harvard University, Cambridge, MA [producer], Ann Arbor, MI: ICPSR [distributor].

King, Gary, Martin Tanner, and Ori Rosen. 1999. "Binomial-beta hierarchical models for ecological inference," *Sociological Methods and Research*, 28, 61–90.

Kleppner, Paul. 1985. *Chicago Divided: The Making of a Black Mayor*. DeKalb, IL: Northern Illinois University Press.

Kullback, S. and R. A. Liebler. 1951. "On Information and Sufficiency." *Annals of Mathematical Statistics*, 22: 525–40.

Lewis, Jeffrey B. 2002. "Comment on McCue, K. F. 2001."The Statistical Foundations of the EI Method' " *The American Statistician*, 55: 106–110.

Morris, Thomas R. and Neil Bradley. 1994. "Virginia." In Chandler Davidson and Bernard Grofman (eds.), *Quiet Revolution in the South: The Impact of the Voting Rights Act, 1965–1990*. Princeton: Princeton University Press.

Rivers, Douglas. 1998. "Review of 'A Solution to the Problem of Ecological Inference.' " *The American Political Science Review*, 92: 442–443.

Robinson, W. S. 1950. "Ecological Correlation and the Behavior of Individuals." *American Sociological Review*, 15: 351–357.

Sabato, Larry. 1981. *Virginia Votes, 1983–1986*. Charlottesville, VA: University of Virginia, Institute of Government.

Sabato, Larry. 1991. *Virginia Votes, 1987–1990*. Charlottesville, VA: University of Virginia, Institute of Government.

Schexnider, Alvin J. 1990. "The Politics of Pragmatism: An Analysis of the 1989 Gubernatorial Election in Virginia," *PS: Political Science and Politics*, 23, 3: 154–156.

Strickland, Ruth Ann and Marcia Lynn Whicker. 1992. "Comparing the Wilder and Gantt Campaigns: A Model for Black Candidate Success in Statewide Elections," *PS: Political Science and Politics* 25, 2: 204–212.

Tate, Katherine. 1994. *From Protest to Politics: The New Black Voters in American Elections*. Cambridge, MA: Harvard University Press.

Traugott, Michael W. and Vincent Price. 1992. "A Review: Exit Polls in the 1989 Virginia Gubernatorial Race: Where Did They Go Wrong," *Public Opinion Quarterly*, 36, 2: 245–253.

Wakefield, Jon C. 2001. "Ecological Inference for 2 × 2 Tables." Typescript Washington University.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57: 348–368.