Understanding King's Ecological Inference Model

A Method-of-Moments Approach

JEFFREY B. LEWIS
Department of Political Science
University of California-Los Angeles

The problem of inferring individual behavior from aggregate data is among the oldest problems in political methodology. Simplest among these problems is the attempt to reconstruct the interior cells of a set of 2 × 2 tables from their marginal totals. An archetypal problem of this sort is the estimation of election turnout by race. In most cases, the number of whites and nonwhites who vote in an election is not tabulated. However, the white and nonwhite population and total turnout in a number of election reporting units (e.g., counties or precincts), is generally known. The problem of ecological inference, in this case (and in all cases considered in this article), is to use partially aggregated information on the marginal distributions of the variables of interest (e.g., percentage white and percentage turnout in each precinct) to infer the joint or conditional distribution of these variables across all reporting units (precincts).

Ecological inference is generally considered to be at best a necessary evil. The glib solution often suggested is "Go collect individual-level data." However, if anything, the need for reliable methods of ecological inference is perhaps greater than it has been at any time since the explosion of survey research in the early 1960s. First, as the tabulation of electoral returns becomes increasingly automated, very large sets of precinct-level election returns are becoming available (e.g., Lewis 1998 and King et al. 1997). A second source of demand is democratization in Eastern Europe and Latin America. Electoral competition in newly formed democracies is an active area of research. However, survey data are limited. Scholars in this field must rely on aggregate election returns to understand these elections (see Ames 1994). Finally, there is demand from an increasing number of scholars interested in quantitative approaches to historical questions. Unfortunately, most existing methods of ecological inference are perhaps most famous for their failures. The lack of robustness of the ecological regression model to violations of its assumptions, for example, represents a real impediment to its application and thus to the use of ecological data in the study of these important substantive problems.

Against this background, Gary King offered *A Solution to the Problem of Ecological Inference* (1997). This new method has received considerable interest and attention. However, its workings and relation to existing methods are largely unexplored. In this article, I attempt to address these questions, and I develop King's model (hereinafter EI) by extending Leo Goodman's (1959) familiar ecological regression (ER) model. I consider analytically a good deal of the distance between what is generally referred to as Goodman's ecological regression and King's estimator.

The empirical application that I use to exemplify the statistical issues inherent in King's method (and ecological inference) is support for ballot propositions. The question considered is how to calculate the percentage support for propositions among those voting for each of two state assembly candidates. I have argued elsewhere that estimates of the support for propositions among those voting for winning assembly candidates can be used as a measure of the policy preferences of assembly members' electoral coalitions (Lewis 1998).

In particular, I consider the problem of using precinct-level voting returns to estimate the support for Proposition 156 among those voting for the Democratic and Republican candidates for California's 60th State Assembly District in 1992. Labeled the Rail Bond Act of 1992, Proposition 156 was the second in a planned series of three rail bonds that were part of a major transportation initiative taken by the state legislature in 1990. The proposition was narrowly defeated statewide by a margin of 49 to 51 percent. The 60th State Assembly District is carved out of Los Angeles County. The district has 301 precincts. Proposition 156 was supported by 40 percent of the 134,000 voters districtwide. The incumbent Republican assembly candidate, Paul Horcher, rather easily defeated his Democratic challenger, Stan Caress, taking 67 percent of the two-party vote.

It should be noted that this proposition-district pairing was not chosen for its substantive interest. Rather, I chose it for two methodological reasons. First, it is one of 306 proposition-district pairs from Los Angeles County for which individual-level voting data are available. Thus, the validity of ecological inferences can be checked against the actual individual-level behavior. Second, of these 306 pairs for which I have individual-level data, this pair is among the most consistent with the assumptions of King's (1997) model. The reader should not therefore infer that the success of King's model in this example is typical of what might be achieved in general. However, an example in which the data are "nice" is preferable for describing how the method works. I go on to test King's method on all 306 data sets for which I can validate its success. From that analysis, one can obtain a good sense of how well King's method works on these sorts of data in general.

The article develops as follows. In the second section, I lay out the notation I used throughout and describe the basic problem of ecological inference. In the next three sections, I provide a method-of-moments logic for understanding King's (1997) model. In the third section, I describe the key assumptions of King's model. In the fourth section, I begin to develop King's estimator from the starting point of the standard ecological regression model. In the fifth section, I complete the development of both King's estimator and my method-of-moments alternative. In the sixth section, I consider King's claim that his method is more robust to violations of its assumptions than the conventional ecological inference model. I conclude with the seventh section.

The Basic Problem and Some Notation

As I mentioned earlier, the problem considered here is that of inferring the distribution of the support for Proposition 156 among voters who supported the Democratic assembly candidate and among voters who supported the Republican assembly candidate. The data used to make these estimates are the marginal distributions of support for the proposition and for the Democratic assembly candidate across precincts. To simplify the problem, I discarded the votes of those who supported third-party assembly candidates or who abstained either in the assembly election or on the proposition. That is, the data fit exactly into the 2×2 table shown in table 1.[1]

I begin with the following definitions:[2]

D_i = vote share for the Democratic assembly candidate (in precinct i),

P_i = vote share for the proposition,

β_i^D = proposition support among Democrats, and

β_i^R = proposition support among Republicans.

D_i and P_i are observed quantities for each precinct $i = 1, \ldots, I$. The object of ecological inference is to make statements about the distribution of β^D and β^R across precincts on the basis of observations on D and P.

One of the contributions of King's book is that it focuses attention on the value of estimating more than just the average or weighted average of the βs across precincts. Indeed, it has often been said that the conventional ecological regression model assumes "constancy" (Freedman et al. 1991, Cho 1998). Constancy is the assumption that $\beta_i^D = \beta^D$ and $\beta_i^R = \beta^R$ for all precincts $i = 1, \ldots, I$, in other words, that the rate of support for the proposition is constant for Democrats and Republicans across precincts. This assumption is obviously rejected in every case and is not an assumption of the regression model proposed by Goodman (1959) (see King [1997, 58–60]). Nevertheless, it is true that ecological regression focuses on the mean to the exclusion of any other feature of the distribution of the βs.

TABLE 1
Inferring Interior Cells of a 2×2 Table from Repeated Observations of Marginals

	Percentage voting Democrat for assembly	Percentage not voting Democrat for assembly	
Percentage voting yes on proposition	β^D ?	β^R ?	P
Percentage not voting yes on proposition	$(1 - \beta^D)$?	$(1 - \beta^R)$?	$(1 - P)$
	D	$(1 - D)$	

Note: Question marks indicate quantities not observed in data. Quantities in interior cells are conditional probabilities.

The focus on the mean or weighted mean of the βs follows from the fact that most often the ultimate quantities of interest are the districtwide values β^D and β^R. In our case, this would be the districtwide percentage of all Democratic and Republican voters who supported Proposition 156. Let these quantities be B^D and B^R, respectively. They are defined as follows:

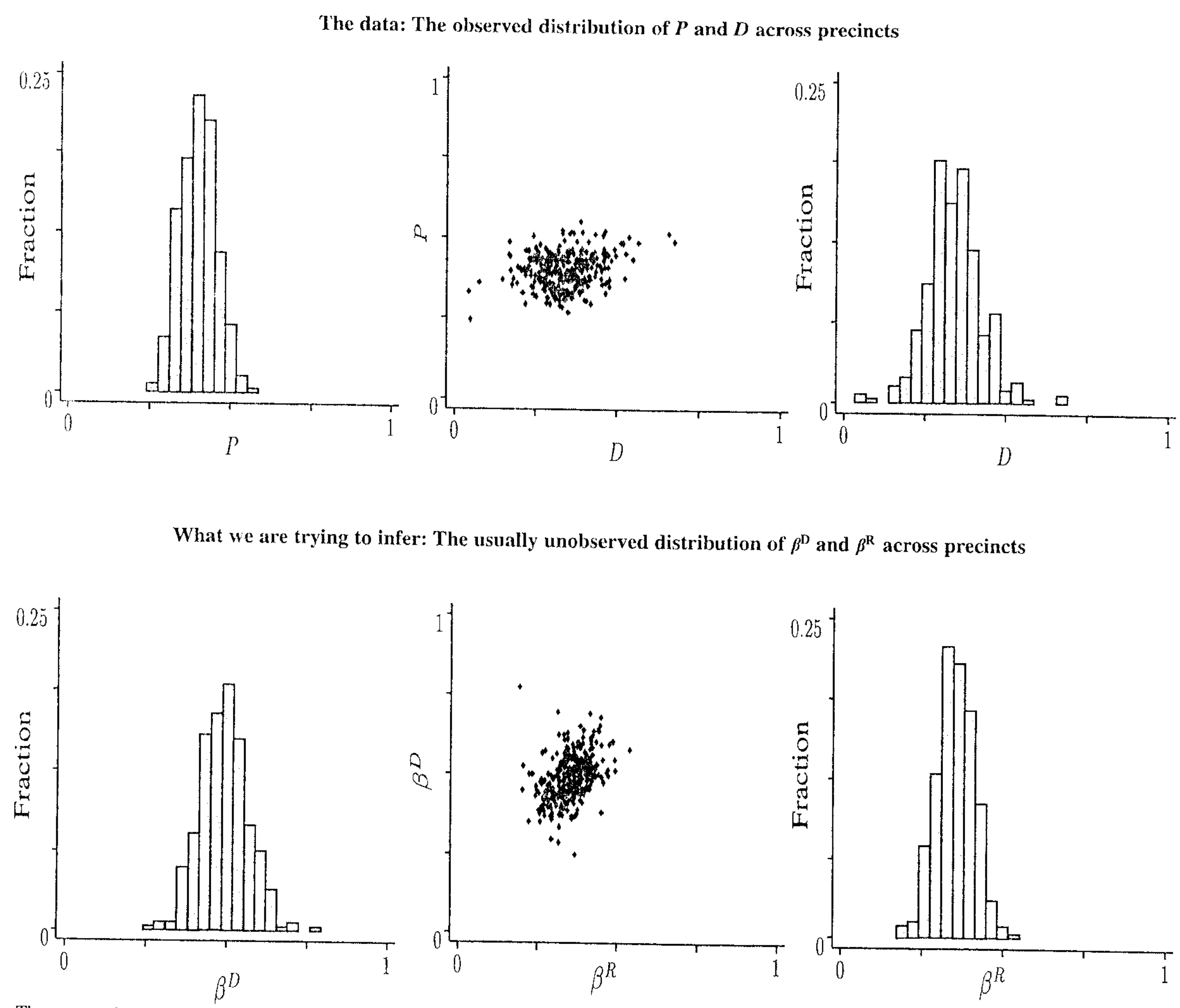
$$B^D = \frac{\sum_i \beta_i^D D_i n_i}{\sum_i D_i n_i} \quad (1)$$

$$B^R = \frac{\sum_i \beta_i^R (1 - D_i) n_i}{\sum_i (1 - D_i) n_i}, \quad (2)$$

where n_i is the number of voters in the ith precinct. In this case, the true $B^D = .49$ and the true $B^R = .35$. That is, the individual-ballot data reveal that 49 percent of Democrats and 35 percent of Republicans supported Proposition 156 districtwide. As King points out, this is not the only quantity that might interest us. In particular, we may want to be able to estimate all of the β^Ds and β^Rs. Or, at the very least, we may be interested in the variance of the βs across districts or the percentage of precincts in which Democrats cast a majority of votes in favor of the proposition.

King's (1997) method seeks to uncover the joint distribution of β^D and β^R. In essence, King proceeds as follows: assume the Ds and Ps observed in the data are random draws from some distribution $g(D, P)$. Then, using these

FIGURE 1
Observed Distribution of P and D across Precincts (the Data)
Usually Unobserved Distribution of β^D and β^R across Precincts (by Attempted Inference)



Note: The top panel shows histograms and a scatter plot of D and P. These are the data. The bottom panel shows histograms and a scatter plot of β^D and β^R. The distributions of β^D and β^R are usually not observed and are in essence the objects that we are interested in determining when making ecological inferences.

observations on D and P, infer the joint distribution of β^D and β^R, $f(\beta^D, \beta^R)$. In our example data, we can directly calculate each β^D and β^R from the individual-level data. Figure 1 shows both the observed distribution of D and P and the usually unobserved distribution of β^D and β^R that we try to infer by using ecological inference. Note that there is considerable variation in β^D and β^R in these data. Democrats gave between 25 and 77 percent support, and Republicans gave between 19 and 53 percent support across precincts. Moreover, there is considerable positive correlation between β^D and β^R. Given estimates of the distribution of the βs, King then generates estimates of each β_i^D and β_i^R conditional on P and D. King accomplishes this by using all of the information that can be gleaned from the data and by making a number of assumptions. The assumptions are covered in the next section.

The structure of the problem leads to the following three observations that have been well understood for some time. As we will see, King's (1997) recombination of these observations is central to his method.

1. *The support of the distribution of* β^D *and* β^R *is* $[0,1] \times [0,1]$.

Because β^D and β^R are conditional probabilities, they must each lie on the unit interval, and therefore each (β^D, β^R) pair must lie on the unit square. More simply, the percentage of Democrats and the percentage of Republicans who supported Proposition 156 must be between 0 and 100. Researchers often use this fact to discredit the standard ecological regression model, which is not subject to this constraint. There are several famous examples in which quantities similar to B^D or B^R have been estimated to be greater than 1 or less than 0.

2. *Conditional on P and D, there is a functional relationship between* B^D and B^R.

In particular,

$$P = \beta^D D + \beta^R (1 - D). \tag{3}$$

Equation 3 is often referred to as the "accounting identity" and is at the heart of both King's method and the conventional ecological regression model. This expression is quite straightforward. The overall support for the proposition is simply the sum of the share of Democrats who supported the proposition times the percentage who are Democrats plus the share of Republicans who supported the proposition times the share who are Republicans. Rearranging this equation, we find that

$$\beta^R = \frac{P}{1-D} - \frac{D}{1-D}\beta^D. \tag{4}$$

Thus, once we condition on P and D, the support of β^D and β^R moves from the entire unit square to a single line through the square. King's rendering of these feasible sets on the unit square is one of the more elegant contributions of his book (King 1997, chap. 5). Figure 2 is an example of these feasible sets for three precincts. As King points out, conditioning estimates of β^D and β^R on P and D greatly reduces our uncertainty. Before we condition, any point in a square is admissible. After conditioning, only those points along a line through the square are admissible.

3. *Facts 1 and 2 place strict bounds on the distribution of* β^D *and* β^R.

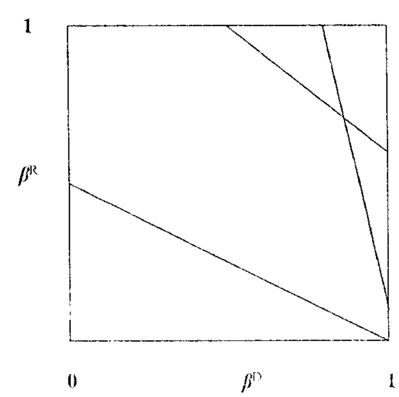
In particular,

$$\beta^D \in \left[\max(0, \frac{P+D-1}{D}), \min(\frac{P}{D}, 1)\right], \quad \text{and}$$

$$\beta^R \in \left[\max(0, \frac{P-D}{1-D}), \min(\frac{P}{1-D}, 1)\right].$$

These bounds have been known in the ecological inference literature since O. D. Duncan and B. Davis (1959) and in statistical theory at least since M. Fréchet (1951). One can find these minimums and maximums by inspecting the points at which the feasible sets shown in figure 2 hit the boundaries of unit square. One of the advantages of King's method is that it forces the estimates of each β_i^D and β_i^R to fall within these logical bounds. The integration of bounding information into Goodman's (1959) regression framework is the single most important innovation of King's solution.

FIGURE 2
Feasible Sets for Three Precincts



Note: Each line represents a single precinct. The points that comprise each line are the set of feasible pairs of β^D and β^R for each precinct given the share of votes for the Democratic assembly candidate and support for the proposition in each precinct.

Assumptions

King's (1997) EI model involves two fundamental assumptions. The first is the assumption that D is uncorrelated with β^D and β^R. The second is that the (β_i^D, β_i^R) pairs are independently and identically drawn from a truncated bivariate normal distribution. The first assumption has a long history in the ecological inference literature. The assumption, its implications, and its history are described in the next subsection. The assumption of the truncated bivariate normal is an innovation of King's that has many desirable properties and a few drawbacks. These are described in a subsequent subsection.

The Independence Assumption: No Aggregation Bias

The structure of the problem alone does not allow us to infer the joint distribution of β^D and β^R from observations on D and P. More must be known or, typically, assumed. In particular, we must know the degree to which β^D and β^R are a function of D. The centrality of this knowledge is cleverly described by the "linear contextual effects model" of Freedman et al. (1991), who propose a world in which $\beta_i^D = \beta_i^R = \alpha_0 + \alpha_1 D_i$. In this world, Republicans and Democrats in each precinct support the proposition at the same level, and that level is an exact linear function of D. Because, by construction, all voters are either Democrats or Republicans, $\beta_i^D = \beta_i^R$ implies $P_i = \beta_i^D = \beta_i^R$, which in turn implies the following exact linear relationship between P and D: $P_i = \alpha_0 + \alpha_1 D_i$. Alternatively, we might be in a world in which the constancy assumption often associated with the Goodman (1959) model holds. In this case, $\beta_i^D = \beta^D$ and $\beta_i^R = \beta^R$ for all i, but β^D need not equal β^R. In this world, we would have (by the accounting identity)

$$P_i = \beta^D D_i + \beta^R (1 - D_i),$$

or, rearranging,

$$P_i = \beta^R + (\beta^D - \beta^R) D_i.$$

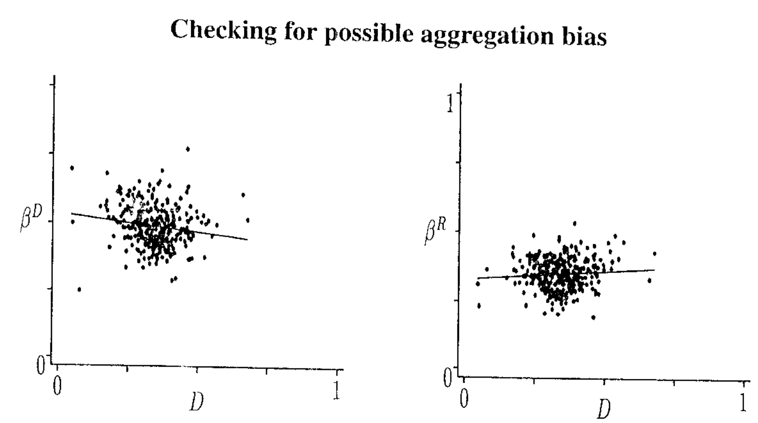
Here, too, we would also expect to see an exact linear relationship between D and P. Thus, without some knowledge of or assumption about the relationship between D and β^D and β^R (such as constancy or linear contextual effects), we cannot pin down the distribution of β^D and β^R from aggregate observations on P and D.[3] When we observe an exact linear relationship between P and D, we cannot tell which of these two models (or infinitely many other models) is correct. This is only the simplest version of the sort of indeterminacy inherent in this problem.

To overcome this indeterminacy, King (1997) follows Goodman (1959) in assuming that β^D and β^R are independent of D. Or, in the parlance of the literature, there is no "aggregation bias."[4] Consideration of aggregation bias can be traced at least to William Robinson (1950).

The problem with the assumption of independence is that it is very hard (if not impossible) to verify empirically. Tests for linear relationships between the βs and D have been proposed by Christopher Achen and W. Phillips Shively (1995), but these tests are relatively weak, as shown by Stephen Ansolabehere and Douglas Rivers (n.d.). Figure 3 shows scatter plots of β^D and β^R against D for our data. Note that there is only a very weak relationship between the βs and D. Thus, in this case, the assumption of independence appears to very nearly hold. As we will see, meeting this assumption is central to the success of the method.

The attraction of this assumption is that, if true, it buys a lot.[5] Goodman (1959) notes that if this assumption holds, one can estimate the means of the βs by simple regression techniques (the ecological regression model described

FIGURE 3
Scatter Plots of β^D and β^R



Note: Only a weak relationship exists between β^D or β^R and D in these data on which King's method should work well. The solid line through the scatter plot is an OLS regression line.

later). Subsequent econometric scholarship has demonstrated that with this assumption, other moments of the joint distribution of βs can be estimated (cf. Hildreth and Houck 1968; Beran and Hall 1992). King found that this assumption along with an assumption about the parametric family of the βs allows the full joint density of the βs to be estimated. Finally, Rivers (1998b) has shown that this assumption alone allows the joint distribution of the βs to be estimated nonparametrically.[6]

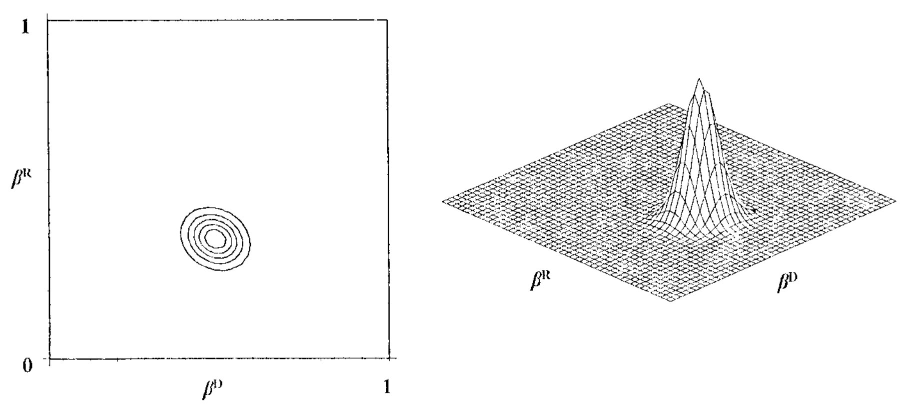
It is important to note that King (1997) makes this assumption.[7] Many scholars have considered aggregation bias (the violation of this assumption) to be the problem in making ecological inferences and therefore may have been misled by King's solution.[8] King's method does not provide a true solution to the problem of aggregation bias. That is, as conventionally discussed and implemented, King's model requires the same conditions as conventional ecological regression to avoid aggregation bias. As is subsequently shown, King's method is less sensitive to the presence of aggregation bias than are other methods. Nevertheless, it is not immune to the problem. Rather, EI should be thought of as a "solution" in the sense that, assuming its assumptions hold, it allows the user to make more efficient estimates B^D and B^R than can be made using conventional regression techniques and also allows the estimation of additional quantities of interest such as the β^Ds and β^Rs.

The Distributional Assumption: The Truncated Bivariate Normal Distribution

Beyond the assumption of independence, King (1997) makes an important assumption about the parametric distribution from which the βs are drawn. The distribution that King chooses is a truncated bivariate normal distribution (TBVN). In particular, King chooses a bivariate normal family in which all the outcomes outside the unit square are truncated. Aside from having support on the unit square, this distribution has a number of desirable properties. Its conditional distributions are simple univariate truncated normal distributions, as is the univariate distribution over any line through the TBVN. Because the feasible sets for the βs are lines, the fact that the distribution over any given line through the truncated bivariate normal is easily derived is quite useful. It greatly simplifies the task of making point estimates of each of the precinct βs. Moreover, the distribution allows for a wide range of empirical situations. The βs can be highly concentrated or relatively dispersed. The distribution can be highly skewed or relatively symmetric. The main restriction that the TBVN puts on the data is unimodality. That is, there must be a most likely pair (β^D, β^R) such that all other pairs are strictly less likely, and each pair of βs that we consider as we move away from this most likely pair (along any line) must be strictly less likely than the last. In essence, there must be only a single mountain, not a whole mountain range. Nevertheless, the TBVN accommodates a wide range of possibilities, requires the estimation of a small number of parameters, and has convenient forms for its conditional distributions. All of these reasons make it a good choice for the application. Figure 4 shows the estimated TBVN density for the proposition voting data. In this case, most of the (β^D, β^R) are estimated to be concentrated relatively close to the mode.

The TBVN distribution is written as follows:

FIGURE 4
Estimated Joint Distribution of β^D and β^R



Note: Two renderings of the same estimated TBVN normal distribution of (β^D, β^R) points in the proposition voting data. The left panel shows the view from the top (iso-density contours). The right panel shows a view from the side (the bivariate density function).

$$f_{tbn}(\beta^D, \beta^R; \mu_D, \mu_R, \sigma_D, \sigma_R, \sigma_{DR}) = \frac{f_{bn}(\beta^D, \beta^R; \mu_D, \mu_R, \sigma_D, \sigma_R, \sigma_{DR})}{\int_0^1 \int_0^1 f_{bn}(x, y, \sigma_D, \sigma_R, \sigma_{DR}) dx dy}. \quad (5)$$

The numerator is the bivariate normal distribution, and the denominator is a normalizing constant (for given parameter values) that forces the distribution to integrate to 1. This expression reveals both the advantages and the disadvantages of the TBVN. On the one hand, because the main component of the distribution is the bivariate normal density function, one can draw on many results in calculating the likelihood of any particular set of values. On the other hand, the denominator is a bivariate normal cumulative distribution function. This integral has no closed form solution and thus must be evaluated by numerical integration.

As I noted earlier, Rivers (1998a) has shown that independence alone is sufficient to estimate the joint distribution of the βs consistently. However, such an approach does not seem particularly practical, because accurate nonparametric estimates of the joint distribution of D and P cannot be made from samples of the size usually associated with ecological inference problems. Moreover, it should be noted that King (1997) can use the assumption of the truncated bivariate normal to relax the independence assumption. King describes extensions to his basic model that allow the βs to be functions of D and other covariates. The identification and tractability of these models depends in large part on the assumption of TBVN. As Rivers (1998b) pointed out, if one assumes that the βs follow an untruncated bivariate normal, King's extended model is not even identified. This lack of identification is problematic because, in many instances (including the example presented here), the degree of truncation is small. In such cases, the TBVN approaches the untruncated bivariate normal and our ability to estimate the relationship between D and the βs is extremely limited. Nevertheless, it is important to note that the imposition of a distributional assumption (if correct) both increases the efficiency of the estimates and may in some situations allow the independence assumption to be relaxed.[9]

Inferring the Distribution of the Unknown Parameters

Given the assumptions that the distribution of the βs is independent of D and that the βs follow a TBVN distribution, the obvious next step is to derive a way of inferring the parameters of that TBVN from observations on P and D. King's (1997) estimator uses the method of maximum likelihood.[10] This method has many advantages. It is almost sure to generate valid estimates of the parameters. Moreover, it will generate consistent and minimum-variance estimates. On the other hand, the method is very computationally expensive and obscures the links between King's method and conventional ecological regression. To highlight the connection with conventional ecological regression and to offer a less computationally burdensome estimator, I now derive a method-of-moments estimator for King's method.

My method-of-moments estimator proceeds in two steps. First, I find five moments of the joint distribution of the βs from observations on P and D. Second, I use these five estimated moments to solve for the parameters of the TBVN.

In developing this estimator, at each stage I compare the estimates of the moments and parameters obtained by my method-of-moments approach and King's ecological estimators with estimates of these quantities made directly from the βs themselves. In this way, the reader will be able to see, at least for this example, how much information is lost when we must estimate these quantities from the ecological data, rather than directly.

A Method-of-Moments Approach

The first order of business is to estimate the means, variances, and covariance of β^D and β^R. The following two propositions adapted from well-known econometric findings about models with random coefficients describe how one can do this using simple regression techniques.

Proposition 1: If β_i^D and β_i^R are independent of D_i and are independently and identically distributed for all precincts $i = 1, 2, \ldots, I$, then consistent estimates of $E(\beta^D) \equiv E^D$ and $E(\beta^R) \equiv E^R$ can be obtained by the ordinary least squares (OLS) regression:

$$P_i = E^D D_i + E^R(1 - D_i) + \varepsilon_i.$$

Let $\hat{E}^D$ be the estimate of $E(\beta^D)$ and $\hat{E}^R$ be the estimate of $E(\beta^R)$.

Proof: See Goodman (1959) or Hildreth and Houck (1968).

Proposition 2: Let $\hat{\varepsilon}_i^2$ be the squared residuals from the regression of P on D. If β_i^D and β_i^R are independent of D_i and are independently and identically distributed for all precincts $i = 1, 2, \ldots, I$, then $V(\beta^D)$ and $V(\beta^R)$ and $Cov(\beta^D, \beta^R)$ can be consistently estimated by the OLS regression:

$$\hat{\varepsilon}_i^2 = V^D D_i^2 + V^R(1 - D_i)^2 + C[2D_i(1 - D_i)] + \gamma_i.$$

Let $\hat{V}^D$ be the estimate of $V(\beta^D)$, $\hat{V}^R$ be the estimate of $V(\beta^R)$, and $\hat{C}$ be the estimate of $Cov(\beta^D, \beta^R)$.

Proof: See Hildreth and Houck (1968) or Judge et al. (1985, chap. 19).

To see the intuition behind these propositions, note that one would like to treat the accounting identity as a regression equation. That is, one would like to estimate the following regression:

$$P_i = \beta_i^D D_i + \beta_i^R(1 - D_i).$$

Of course, as a regression, this equation looks a bit strange. First, its parameters β_i^D and β_i^R vary across observations. Second, the equation has no error term. The problem of regression models with randomly varying coefficients has been widely treated in the econometrics literature (cf. Hildreth and Houck 1968; Griffiths 1972; and Judge et al. 1985, chap. 19). The solution is to rewrite the identity as

$$P_i = (E^D + \varepsilon_i^D)D_i + (E^R + \varepsilon_i^R)(1 - D_i),$$

where E^D and E^R are the means of β^D and β^R, respectively, and $\varepsilon_i^D = \beta_i^D - E^D$ and $\varepsilon_i^R = \beta_i^R - E^R$. Rearranging the above equation, we find

$$P_i = E^D D_i + E^R(1 - D_i) + \varepsilon_i^D D_i + \varepsilon_i^R(1 - D_i)$$
$$= E^D D_i + E^R(1 - D_i) + \varepsilon_i$$

or

$$P_i = E^R + (E^D - E)D_i + \varepsilon_i, \qquad (6)$$

where $\varepsilon_i = \varepsilon_i^D D_i + \varepsilon_i^R(1 - D_i)$. Equation (6) looks much more like a conventional regression having fixed parameters and an error term. Fitting equation (6) is what is conventionally referred to as Goodman's (1959) ecological regression. In Goodman's model, estimates of E^D and E^R are taken as estimates of B^D and B^R. $\hat{E}^D$ and $\hat{E}^R$ will be consistent estimators for B^D and B^R so long as the total number of voters in each precinct is uncorrelated with $\beta^D + \beta^R$ (see King 1997, 61).[11] Proposition 1 asserts that this equation can be estimated consistently by OLS regression. However, as Goodman (1959) himself noted, the residual in equation (6) is heteroskedastic. In particular,

$$V(\varepsilon_i) = V(\varepsilon^D)D_i^2 + V(\varepsilon^R)(1 - D_i)^2 + 2\text{Cov}(\varepsilon_i^D, \varepsilon_i^R)D_i(1 - D_i).$$

Thus, one can achieve more efficient estimates of E^D and E^R by using a feasible weighted least-squares estimator that uses square roots of the predicted values of the regression in proposition 2 as weights.[12] In either case, the notion is that one can use simple regression techniques to estimate the means of the unobserved β^D and β^R. It should be pointed out that this technique does not incorporate the logical bounds on the E^D and E^R. That is, although E^D and E^R must both lie on the unit interval, there is no guarantee that $\hat{E}^D$ and $\hat{E}^R$ will do so (in finite samples). Even if the assumptions of the model hold, we may still obtain infeasible estimates. However, unless the sample size is very small or E^D and E^R are very close to 0 or 1, this should not be a problem. Estimates of $\hat{E}^D$ and $\hat{E}^R$ that lie well outside the unit interval are more likely the result of aggregation bias.[13]

According to proposition 2, beyond the means of the βs, we can also estimate the variances and covariance. Some intuition as to why this is true comes simply from inspecting the expression for the variance of ε given above and noting that $E(\varepsilon^2) = V(\varepsilon)$. Again, there is no guarantee that the OLS estimates of the variances and covariances will be feasible. That is, the estimated variances may be negative or the estimated covariance and variances may imply a correlation between β^D and β^R that is greater than 1 or less than −1. In these cases, one can use constrained regression techniques to obtain feasible estimates. Such estimators have been suggested in the literature (e.g., Schwallie 1982). Although infeasible variance and covariance estimates may indicate a violation of the independence assumption, they often occur simply by chance because of the relatively large degree of uncertainty that is often involved in their estimation.

Together, propositions 1 and 2 suggest a two-step technique for inferring the means, variances, and covariance of β^D and β^R. First, regress P on D, then regress the squared residual from the first regression on D_i^2, $(1 - D_i)^2$, and $2D_i(1 - D_i)$. One can then achieve more efficient estimates by using predicted values from the second regression to weight the first regression. Also, if need be, one can use constrained regression techniques to obtain feasible estimates.

Table 2 contains the estimated means, variances, and covariance of β^D and β^R. The second and third columns show the ecological estimates made by the procedure previously described. The fourth and fifth columns show estimates made from the usually unobserved βs themselves. These "direct estimates" are just ordinary summary statistics for the proportion of Democrats and the proportion of Republicans who supported Proposition 156 across 301 precincts.[14] Here we see that with near-independence between D and the βs and 301 precincts, ecological techniques very accurately recover the means of β^D and β^R. The ecological estimates are nearly identical to those made from the observations on β^D and β^R directly. There is, however, evidence of information loss. The standard errors from the direct estimates are four to five times smaller than the standard errors from the ecological estimates.

There is somewhat less information in the ecological data for estimating the covariance and variances. We see larger differences between the ecological and the direct estimates. We also see estimated standard errors that are ten to fifteen times larger for the ecological estimates. Thus, even when

TABLE 2
A Few Estimated Moments of the Joint Distribution of β^D and β^R across Precincts: Estimated Means, Variances, and Covariance of β^D and β^R

	Ecological estimate		Direct estimate	
Parameter	Estimate	*SE*	Estimate	*SE*
$E(\beta^D)$	.4911	.0244	.4911	.0043
$E(\beta^R)$	.3496	.0127	.3518	.0033
$V(\beta^D)$	.0042	.0064	.0057	.0005
$V(\beta^R)$	.0031	.0019	.0033	.0002
$\text{Cov}(\beta^D, \beta^R)$	.0021	.0035	.0017	.0003
$\text{Corr}(\beta^D, \beta^R)$	.61		.40	
N	301		301	

the independence assumption is very nearly met and the sample size is relatively large, we have only a limited ability to recover the variances and covariance of the βs. Nevertheless, the estimates are fairly reasonable. The variance of β^D is found to be larger than that of β^R in each case, and the estimated covariance is similar in magnitude.

Overall, when the independence assumption is met, the regression technique developed above does a rather good job of estimating the means, variances, and covariance of the two unobserved variables. Although there is a fair amount of lost information, it is nevertheless impressive, and indeed somewhat comforting, that so much information can be culled from the aggregate data.

Moving from the Means and Variances to the Parameters of the TBVN

As noted earlier, one of the main contributions of King's method is that it goes beyond means and variances to fully describe the joint distribution of the βs. This is accomplished through the assumption that the joint distribution of the βs is truncated bivariate normal. What we now need is a mapping from the five estimated moments (means, variances, and covariance) of β^D and β^R to the parameters of a truncated bivariate normal. Solving for the parameters of the TBVN in this way is by no means algebraically straightforward. S. M. Shah and N. T. Parikh (1964) give equations for the mapping between the parameters of the TBVN and moments estimated above,

$$T: (\mu_D, \mu_R, \sigma_D, \sigma_R, \sigma_{DR}) \rightarrow (E^D, E^R, V^D, V^R, C),$$

where T is a mapping (set of functions) that describes each mean, variance, and covariance as a function of the five parameters of the distribution. Because there were a few typesetting errors in Shah and Parikh's article, I give these equations in appendix A. Plugging in the estimated moments found in the previous section, I have a system of five equations (one for each moment) and five unknowns (one for each parameter of the distribution). The solution to this system of equations cannot be achieved analytically. That is, there is no way to express the five parameters as a function of the five moments. However, it is possible to solve the system numerically. Formally, the problem is to find

$$T^{-1}: (E^D, E^R, V^D, V^R, C) \rightarrow (\mu_D, \mu_R, \sigma_D, \sigma_R, \sigma_{DR}).$$

I do this by using Newton's method, which involves an iterative series of numerical approximations that converge to the solution. Although this method involves a considerable amount of computation, it is much less computationally intensive than the estimator for the TBVN parameters that King (1997) employs and that is described below.

Table 3 contains the method-of-moments and maximum-likelihood estimates of the parameters of the distribution of the βs made from direct and ecological observations. Note

TABLE 3
Estimated Parameters of the Truncated Bivariate Normal (TBVN)

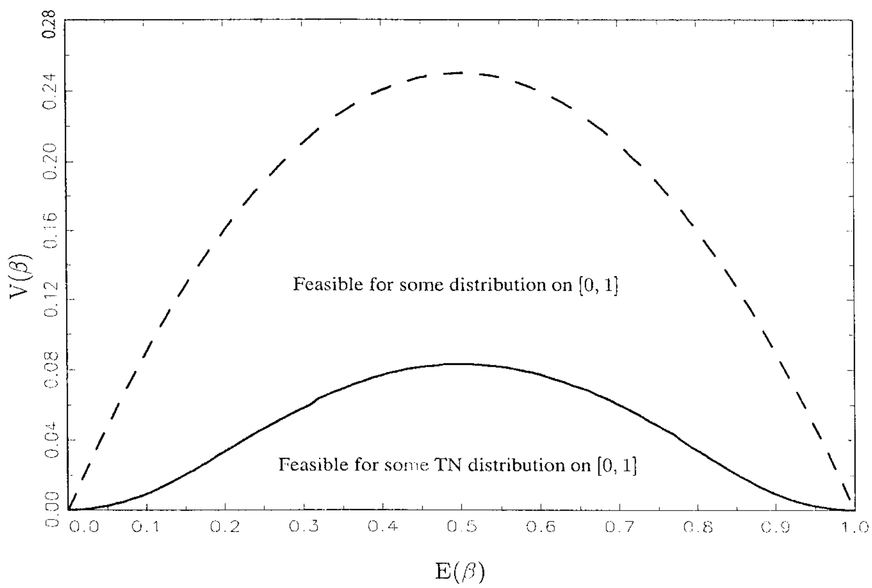
Parameter	MOM Estimate	MOM *SE*	ML Estimate	ML *SE*
	Ecological estimate			
μ_D	0.4912	0.0244	0.4835	0.0309
μ_R	0.3496	0.0127	0.3532	0.0159
σ_D	0.0649	0.0500	0.0886	0.0364
σ_R	0.0559	0.0166	0.0615	0.0109
σ_{DR}	0.0021	0.0035	0.0005	0.0028
ρ	0.6063		0.0925	
	Direct estimate			
μ_D	0.4911	0.0043	0.4911	0.0043
μ_R	0.3518	0.0033	0.3518	0.0033
σ_D	0.0754	0.0225	0.0754	0.0879
σ_R	0.0573	0.0201	0.0573	0.0863
σ_{DR}	0.0017	0.0003	0.0017	0.0621
ρ	0.4020		0.4020	

Note: Estimates of the parameters of TBVN were made both by maximum likelihood (ML) and by the method of moments (MOM) using alternatively direct observations on β^D and β^R and ecological observations of P and D. The standard errors estimated for the MOM estimators were calculated using the "delta" method.

that because there is very little truncation in our example, the estimated parameters of the truncated bivariate normal are very close to the estimated moments.[15] Indeed, the estimates of μ_D and μ_R are nearly identical to the estimates of E^D and E^R. Again, in this case, there is relatively little difference between the direct and indirect estimates. In general, the estimated standard errors are larger in the ecological case (as we would expect). Most interesting, the estimated standard errors for σ_D and σ_R are larger for the direct estimates than they are for the ecological estimates in which maximum likelihood was used.[16]

One problem this method faces is that the TBVN places additional restrictions on the feasible values of the means and variance beyond the ones applied above. For example, because the TBVN is unimodal, not only does the variance of each of the βs have to be positive but it must also be less than 1/12—the variance of the least unimodal (i.e., a uniform) distribution on the [0,1] interval. Moreover, as the variance approaches 1/12, the mean must approach 0.5. As the variances β^D and β^R both go to 1/12, their absolute correlation goes to 0 or 1. Figure 5 contains an example of these constraints in the univariate case. The figure shows the maximum possible variances of a truncated normal random variable β with a given mean as well as the maximum possible variance that any random variable on the [0,1] interval can have for a given mean. There is considerable room for the means and variances to fall outside those feasible for the TBVN, even when the independence assumption holds. These are the simplest restrictions. More gener-

FIGURE 5
Feasible Mean Variance Pairs for Distributions on the [0,1] Interval



Note: All of the points under the dashed curve are feasible mean and covariance combinations for a distribution with support on the [0,1] interval. All of the points under the solid curve are feasible mean and variance combinations for some truncated normal (*TN*) distribution. The maximum of the dashed curve is 1/4, the variance of a Bernoulli distribution. The maximum of the solid curve is 1/12, the variance of a uniform distribution.

al restrictions limit the maximum variance and correlation that can hold for a given set of means, or vice versa. If the βs do in fact follow the TBVN and the number of observations is reasonably large, these restrictions present few problems. For most real-world data, one must put constraints on the estimated moments to ensure that the inverse moment calculation is possible. However, these moment restrictions allow the TBVN assumption to be tested in the data. On the other hand, because the exact form of the moments restrictions is complicated, a maximum-likelihood method that directly estimates the parameters of the TBVN (such as King's 1997 estimator) may be preferable.

A Maximum-Likelihood Approach

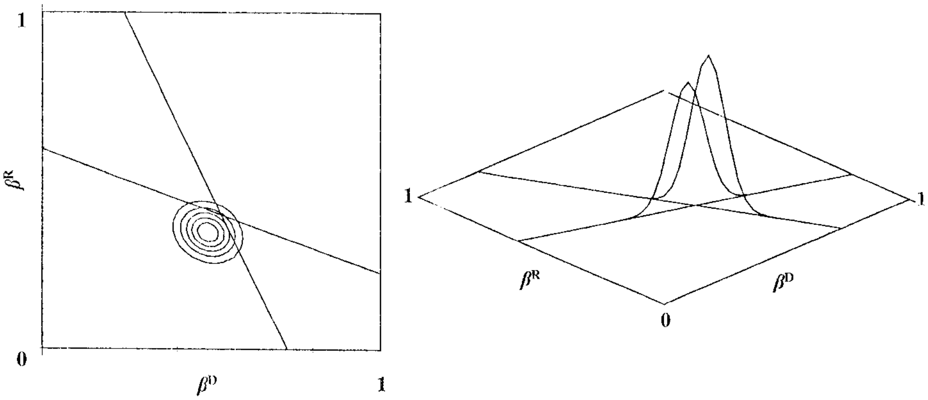
King's (1997) method estimates the parameters of the TBVN distribution of the βs directly and simultaneously by maximum likelihood.[17] I present a derivation of the likelihood function in appendix B. The function involves two additive terms. One term is a weighted least-squares regression minimand that is similar to the weighted regression for estimating the means of the βs given in an earlier section describing the method-of-moments approach. The other, a more complicated expression, accounts for the fact that the βs must fall within their logical bounds. This likelihood function must be maximized numerically and involves the repeated calculation of bivariate and univariate normal cumulative functions. Although, in theory, the maximum-likelihood estimator is clearly superior to the method-of-moments estimator, inaccuracy in the calculation of the normal integrals combined with the inherent difficulty involved in estimating the σs described earlier sometimes leads to convergence problems and slow performance in this estimator. Given the estimated μs and σs, King's procedure for calculating the βs and Bs is basically equivalent to the procedure described next.

Estimating the Precinct-Level Quantities of Interest

Having described a method of estimating the moments of the distribution of the βs and from those the parameters of their assumed underlying TBVN distribution, I now turn to how one can use these parameters to estimate each β_i^D and β_i^R. Generating estimates of each β_i^D and β_i^R conditional on P_i and D_i is a major innovation of King's (1997) method. Armed with these estimates, one can then calculate B^D and B^R or any other quantity that might be of interest. A full treatment of how to obtain estimates of each β_i^D and β_i^R is given in appendix C. Here, I give a graphical intuition for the method by which this is accomplished. The way these estimates are derived highlights the ways EI uses all of the information contained in the structure of the problem.

The estimated parameters of the TBVN define a distribution of (β^D, β^R) pairs over the unit square. This distribution

FIGURE 6
Estimated Distribution of β^D and β^R Conditional on P and D for Two Precincts



Note: These graphs place the feasible (β^D, β^R) pairs for two of the 60th State Assembly District's precincts onto the graphs in figure 4. The right-hand panel demonstrates one of the nice properties of the TBVN distribution. The distribution of points over every precinct feasible set (line through the graph) is a truncated univariate normal distribution. This fact greatly facilitates the calculation of precinct-level quantities of interest.

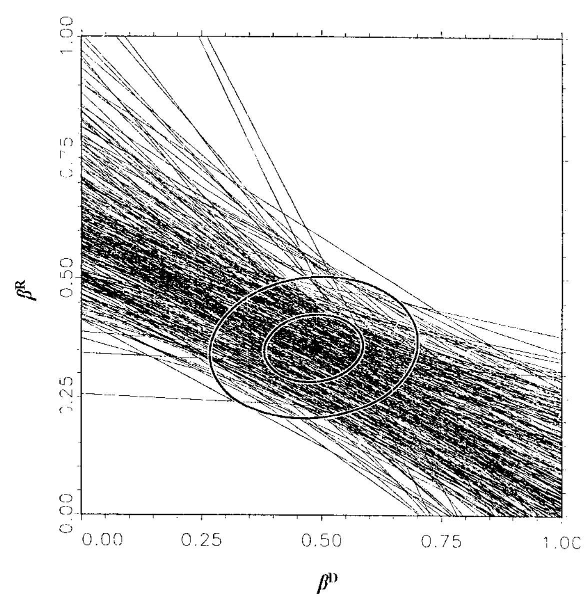
is shown in figure 4.[18] Conditional on P and D, the feasible (β^D, β^R) pairs lie along a single line through the unit square (see figure 2). The trick to estimating the precinct quantities is to compute the mean of the distribution of β^D and β^R over each of these feasible sets.

Figure 6 shows the distribution of (β^D, β^R) pairs over the feasible sets for 2 of the 301 precincts. Note that the distribution over each of the precinct lines appears to be (truncated) normal—and indeed it is. This is a useful feature of the TBVN distribution. Because the distribution over the feasible set is normal, it is quite straightforward to find the expectation (mean) over the line using well-known formulas for the expectations of truncated normal variables. These estimated expected values serve as our estimates of each β^D and β^R.

Figure 7 contains the feasible sets for all 301 precincts in the data. One then forms estimates of the β^Ds and β^Rs by taking the expectations over each of these feasible sets. Figure 8 graphs the predicted βs against their true counterparts. The method-of-moments and King's (1997) estimator produce very similar estimates (most of the points in the top two panels of the figure are close to the 45° line). Both methods produce estimates that are reasonably close to the observed values.[19] In table 4, the estimates of the district-level quantities of interest B^D and B^R are compared. Note that all the methods produced quite good results in this case.

Although the point of this example was not to demonstrate the effectiveness of King's (1997) approach (whether estimated by maximum likelihood or by the method of moments), the example does demonstrate that the approach can work very well when its assumptions are met. It is useful to note that because these assumptions are basically the same as those made by Goodman's (1959) ecological

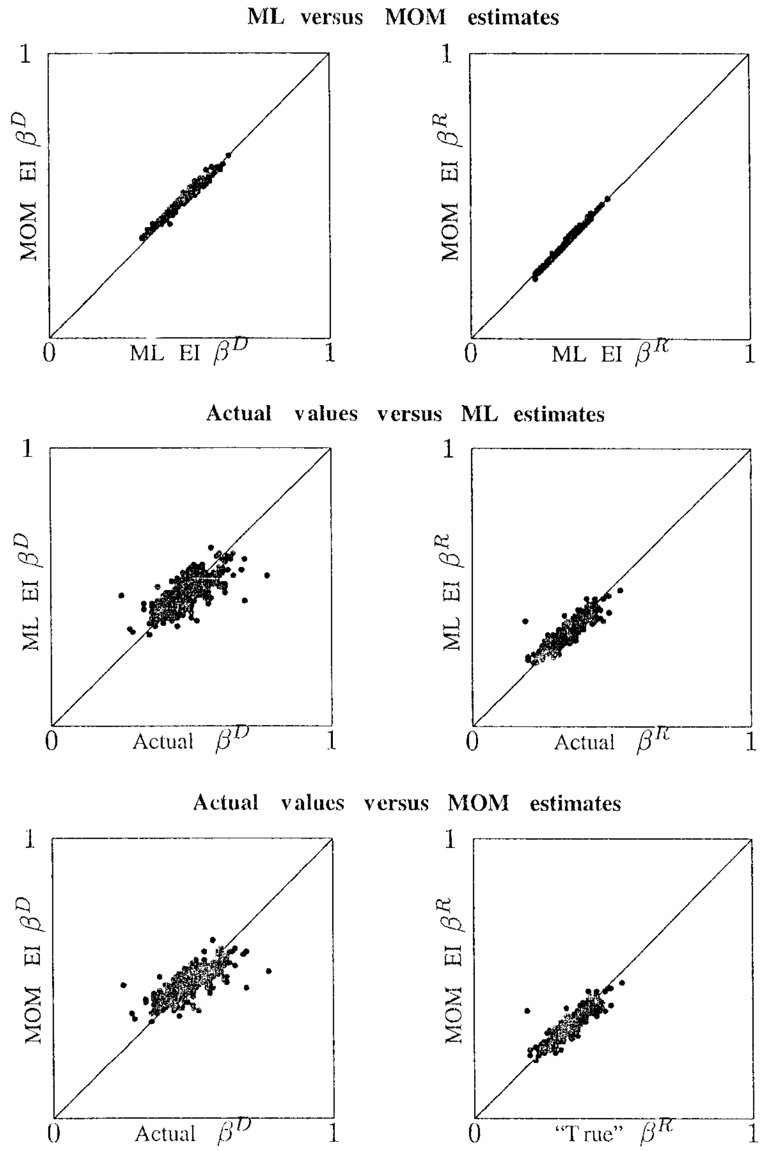
FIGURE 7
King's Tomography Plot



Note: Each line represents the set of feasible (β^D, β^R) pairs for one of the 301 precincts in the 60th State Assembly District. The circles are iso-density contours of the TBVN distribution. The outer contour circumscribes 80 percent of the density of the distribution.

regression, it also works well in this case. Although this might suggest that one should simply use ecological regression, such a conclusion would be unwarranted. First, King's method allows estimation of quantities other than B^D and B^R (the districtwide proportions). Second, King's method restricts all (β^D, β^R) pairs so that they fall on their feasible

FIGURE 8
Predicted βs against Their True Counterparts



Note: Plots the distributions of the precinct-level estimates of the percentage of Democrats (β^D) and the percentage of Republicans (β^R) that supported Proposition 156 against each other and against the "true" values obtained from individual-level data. The plots include both the method-of-moments estimates (MOM) developed here and estimates using King's maximum-likelihood (ML) estimator.

sets. Thus, any quantities estimated by King's approach are guaranteed to be, at the very least, feasible. This is not true of ecological regression, even if its assumptions are met. Indeed, assuming that the independence assumption is met, it requires some effort to concoct a circumstance in which one would not want to use King's method over the traditional ecological regression model.

Aside from what are basically efficiency improvements, King (1997) also claims that his model is more robust to violations of the independence assumption. This claim of robustness comes in two forms. King notes that his model can be extended in ways that explicitly incorporate dependence between the Ds and the βs, as was discussed briefly earlier. Beyond this, King claims that his model is relatively robust to violations of the independence assumptions. This claim has proven more controversial (see Cho 1998 or Freedman, Klein, and Ostland 1998 and responses). I consider this claim in the next section.

TABLE 4
District-Level Estimates of the Percentage of Democrats and Republicans Supporting Proposition 156, by Various Methods

Quantity	Goodman's ER estimate	MOM EI estimate	EI estimate	Actual value
Proposition support among				
Democrats	0.492	0.488	0.481	0.485
Republicans	0.349	0.346	0.346	0.347

Note: All three methods produced results very close to the true values. ER = ecological regression; MOM = method of moments; EI = ecological inference.

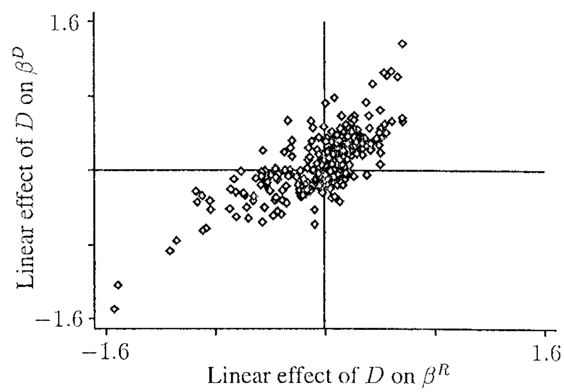
Violating the Independence Assumption

In this section, I consider the degree to which King's method is robust to violations of the independence assumption. For this inquiry, I use 306 proposition–assembly district pairs for which I have individual-level data. In particular, I make ecological inferences about the percentage of Democratic and Republican state assembly voters who supported each of 13 propositions in each of 25 districts. For each of these pairs,[20] I compare the estimates of B^D and B^R to their true (realized) values.

I make the ecological estimates by four methods. Three of these—King's (1997) maximum-likelihood estimator, my method-of-moments estimator, and Goodman's ecological regression (ER)—are by now familiar. In addition, to these is a fourth estimator that I call fastEI, which is basically the same as the method-of-moments estimator except that rather than using T^{-1} to find the parameters of the TBVN, it simply equates the moments and the corresponding parameters of the TBVN. The method is fast because it skips this computationally difficult step. Of course, the estimates of the TBVN parameters made in this way are biased and inconsistent. However, if the degree of truncation is small, as was the case above, the size of the bias will be small. In any event, fastEI is a simple and easy-to-implement routine that involves no numerical maximization or evaluations of the bivariate normal cumulative distribution function. Because of its simplicity, it provides a quick and easy way to approximate the results that we would find using King's method.[21] Moreover, fastEI allows us to see how important getting the parameters of the TBVN "right" is to the success of King's model. That is, if most of the advantage of King's model comes from constraining the βs to fall within their feasible sets and not on precisely estimating the parameters of the TBVN, then the performance of fastEI should be similar to that of King's maximum-likelihood estimator.[22]

Figure 9 plots the slope coefficients of regressions of the true β^D and true β^R for each of the 306 proposition–assembly district pairs. Note how far many of these points fall from the origin (the point at which neither β^D nor β^R is a linear function of D). In other words, among these data sets, we find many that grossly violate the key independence assumption of King's method as well as others that only modestly violate this key assumption. Thus, these data present a good test of the robustness of the various estimators to aggregation bias.

FIGURE 9
Slopes of Regressions β^D and β^R on D across 306 Proposition–Assembly District Pairs



Note: Graph shows slope coefficients for regressions of β^D and β^R on D for each proposition–assembly district pair. Each point represents one proposition–assembly district pair. Note that many of the points fall far from the origin (the point at which there would be no linear relationship between β^D or β^R and D).

Table 5 contains the mean and maximum absolute deviation between the true and estimated values of B^D and B^R found by using each of these four methods. The ER behaves as expected in the presence of aggregation bias (the violation of the assumption that D and the βs are independent). The average differences between the Goodman estimates and the truth are larger than under the other methods, and the maximum differences are quite far from the truth. The maximum deviation for B^R demonstrates the infeasible estimate problem inherent in ER. The maximum difference of 1.073 could have been produced only by an estimated B^R

TABLE 5
Deviations of District-Level Quantities of Interest from Actual Values

	Mean absolute deviation from truth		Max. absolute deviation from truth	
Method	B^D	B^R	B^D	B^R
ER	.092	.123	.396	1.073
MOM EI	.108	.122	.678	.756
FastEI	.078	.099	.310	.523
ML EI	.080	.099	.322	.487

Note: Average absolute "error" and maximum absolute "error" of estimates of the percentage of Democrats (B^D) and Republicans (B^R) supporting Proposition 156 for 306 proposition–assembly district pairs, by four methods of ecological inference.

greater than 1 or less than 0. Averaging across the two district-level quantities of interest and the 306 data sets, errors made by King's (1997) method were 16 percent smaller than those made by ER.

The method-of-moments estimator avoids the problem of logically impossible parameter estimates but otherwise performs no better than Goodman's ER. Because the estimators of the moments can be badly biased when the independence assumption is violated and the estimates of the parameters of the TBVN from the moments are very sensitive to the estimates of the moments, this result should not be surprising. In many instances, I had to constrain the estimated moments to calculate the parameters of the TBVN at all. All of these problems limit the feasibility of the full method-of-moments estimator when the assumptions of the model (either independence assumptions or distributional assumptions) are violated.

However, the fastEI estimator is more effective. Using the moments themselves as estimators for the parameters of the TBVN and calculating the precinct βs in the same way as EI, leads to substantial improvement over ER. The maximum-estimation errors are greatly reduced compared with their values when ER is used. The fastEI estimator improves the prediction by 2 percentage points on average over the Goodman (1959) regression.

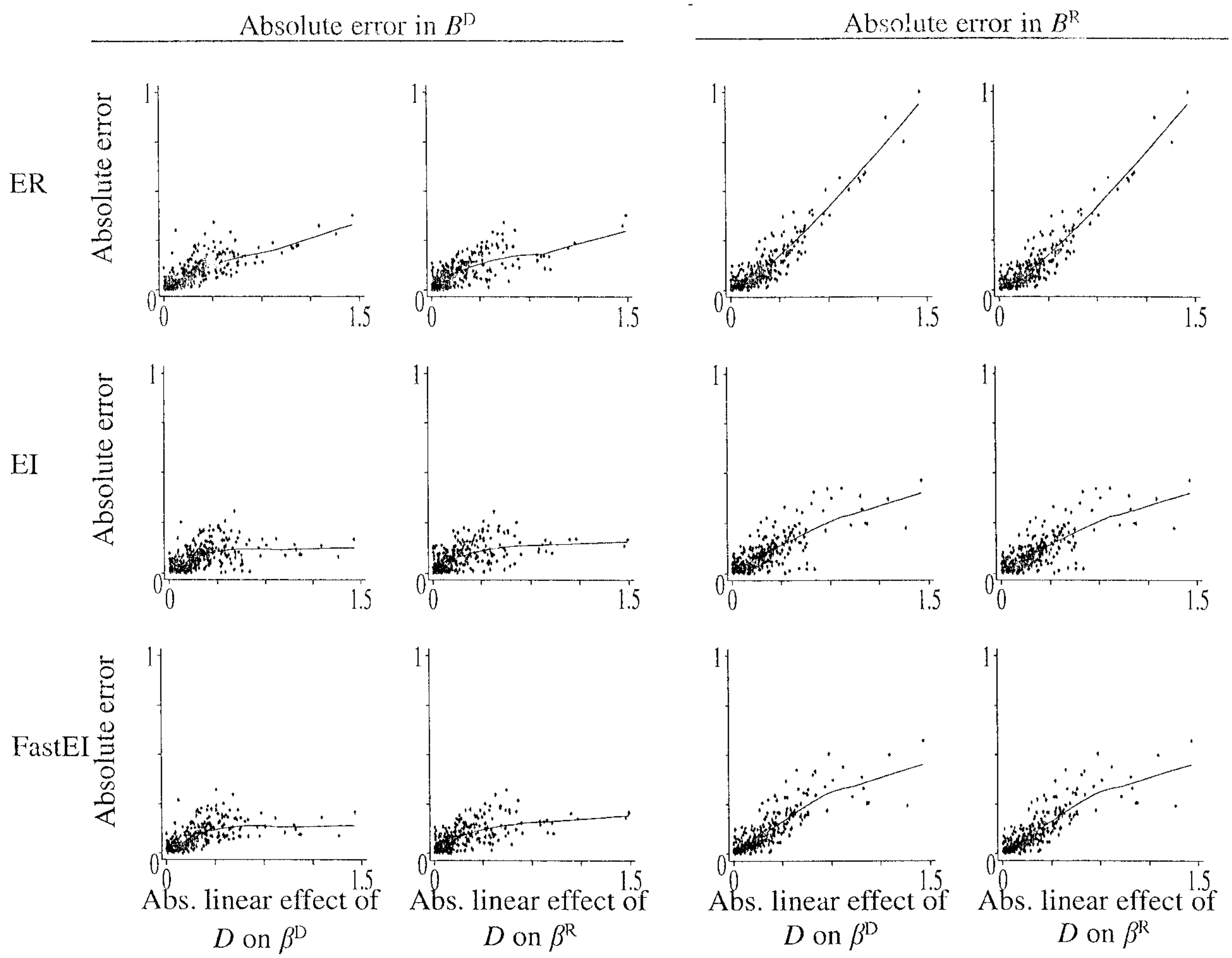
The performance of King's (1997) estimator is very similar to that of fastEI. By using all available information in the calculation of the parameters of the TBVN (and by including priors on the parameters), EI outperforms the method-of-moments estimators on these data. Even though it no doubt produces better estimates of the parameters of the TBVN, EI offers no improvement over the fastEI procedure in these data. Given that the distributional and independence assumptions are strongly rejected in many of the data sets, this is perhaps not surprising.

Faced with a set of problems in which the independence assumption is often violated, all of the methods considered made significant errors. King's (1997) method provided better estimates than the standard ecological regression model. However, none of these methods was particularly robust, making average errors of at least 8 percentage points and maximum errors of at least 30 percentage points.

Table 5 does not directly address the question of whether King's method is more robust than ecological regression. Figure 10 addresses this question. For each of the 306 data sets, I ran regressions of the actual β^D and β^R on D. The estimated absolute slope coefficients for these regressions are plotted in figure 9. The scatter plots in figure 10 plot these same estimated slope coefficients against the absolute error of the estimated district-level quantity of interest (B^D or B^R) for each of three ecological inference methods. In every case, as the degree of violation of the assumption that D and a given β increases (larger slope coefficient), the average absolute error of the prediction increases. The prediction line through each graph is a locally weighted least-squares (LOWESS) regression line. This LOWESS line allows for nonlinear data fitting and is quite revealing in this case. Note that from all three methods, when the independence assumption is met (values of the absolute slope coefficient near 0), the prediction errors are quite small. Furthermore, as the degree of violation of the independence assumption increases from 0, the predicted absolute errors also increase. However, the rate at which the estimation errors increases as a function of the degree of nonindependence differs greatly among the methods and between the estimates of the Democratic and Republican support for the propositions. For each method, the estimate of Republican support is generally more strongly affected than is the estimate of Democratic support. This result is a consequence of the fact that Republicans are the minority of voters in most of the Los Angeles County assembly districts. Because Republicans are the smaller group, when a given number of proposition voters are incorrectly attributed to the wrong party, the misattribution has a larger effect on the estimated fraction of the Republicans who support the proposition than it does on the estimated fraction of Democrats who support the proposition.[23] Comparing across methods, we found that EI and fastEI do in fact demonstrate some robustness to violations of the independence assumption. Although the absolute errors were increasing in the degree of dependence of the βs on D for all methods, the errors made by EI and fastEI were much less responsive to the degree of dependence. Indeed, we see that past a certain level of dependence, the errors made by EI and fastEI hardly increased at all. The explanation is that the logical bounds the methods impose on the estimates in some sense "cap" the size of the potential error.[24] Note also that there are very important cross-effects. That is, a violation in the assumption that β^D and D are independent has important consequences not only for the estimation of B^D but also for the estimation of B^R.[25] This is true for ecological regression as well as for the other methods.

Overall, King's (1997) method appears to be more robust to violations of the independence assumption than Good-

FIGURE 10
Effect of Aggregation Bias on Estimates of the District-Level Quantities of Interest



Note: Figure shows the relationship between absolute prediction errors in the districtwide shares of Democrats and Republicans supporting each proposition and the degree of violation of the assumption that D and the βs are independent as measured by the estimated linear effect of D on each β. The best fit lines shown in each panel are robust locally weighted regression lines.

man's ecological regression in the sense that King's method produces superior estimates when the independence assumption is violated. At the district level, EI improved by 16 percent the predictions of ecological regression. However, in an absolute sense, King's method (at least in its basic form) is not robust to violations of the independence assumption. That is, when applied to series of data sets that substantially and increasingly violated the independence assumption, the EI estimates substantially and increasingly missed their mark.

Conclusion

King's (1997) model represents a real advance in the technology of ecological inference. As demonstrated above, it is an advance that builds directly on Goodman's (1959) pioneering work. The method-of-moments estimator developed here highlights this connection. In essence, King's method can be thought of as a truncated random-coefficients version of Goodman's ecological regression. By including information about the logical bounds of the parameters, King's method improves the efficiency of the estimates of the district-level quantities of interest and allows the user to make inferences about the precinct-level quantities. Indeed, King's method allows the analyst to make myriad statements about the distribution of precinct quantities. Before King, no such technology was readily available.[26] In this sense, King's estimator is a solution to important problems of ecological inference.

On the other hand, in its basic and most common formulation, King's (1997) method makes the same independence assumptions that are used in ecological regression. Concern about the empirical validity of this assumption has been expressed since Robinson (1950) and Goodman (1959). Indeed, this concern has been so central to the literature that it is commonly thought of as *the* problem of ecological inference. In this sense, King's model is not (in general) a solution. Although King's model is more robust to aggregation bias than conventional regression techniques, it is not robust in an absolute sense. That is, violations of the independence assumption will in general lead to estimated quantities that are far from the truth.

With respect to the problem of aggregation bias, King's method may provide a partial solution in two ways. First, extensions to King's method do indeed allow analysts to relax the independence assumption. Second, King offers some diagnostic tools to detect situations in which the independence assumption has been violated. A careful treatment of these contributions is beyond the scope of this article. Readers should consult Cho (1998) or Freedman, Klein, and Ostland (1998) for a consideration of these features.

This study is meant to give the reader a better sense of King's method and its connection to the past. The main conclusion is that EI offers a good technology for ecological inference for cases in which the independence assumption is not grossly violated. It is not, however, a panacea. In the end, we must remember that aggregation involves the loss of information and that no statistical method can recover what is lost. Thus, as with all statistical methods, King's EI must be used with caution and with careful consideration of what it requires and how it works.

Keywords: ecological inference, random coefficients, aggregate voting data, California assembly elections

APPENDIX A
Moments of the TBVN

Recursion equations for, and the first two moments of, the doubly truncated bivariate normal distribution are given by Shah and Parikh (1964). The equations they present include several small errors. Because I know of nowhere in which these equations appear without error in the literature, I present them here following Shah and Parikh's (1964) notation. If x and y are drawn from the double truncated standard bivariate normal distribution, their moments are given by:

$$E(x^r, y^s) = \int_h^l \int_k^m \frac{x^r y^s e^{-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}}}{2\pi\sqrt{1-\rho^2}\, R(l,h,m,k,\rho)}\, dy dx,$$

where h and k are the lower truncation points on x and y, l and m are the upper truncation points on x and y, ρ is the correlation between x and y, and $R(l, h, m, k, \rho)$ is the value of the untruncated bivariate normal cumulative distribution over the domain of x and y. Letting $z(x)$ be the standard normal density function and

$$\varphi(x) = \int_x^\infty z(s)ds,$$

the first 2 moments of the joint distribution can be written as:

$$R \cdot E(x) = z(h)\left(\varphi(\frac{k-h\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-h\rho}{\sqrt{1-\rho^2}})\right) - z(l)\left(\varphi(\frac{k-l\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-l\rho}{\sqrt{1-\rho^2}})\right)$$
$$+ \rho z(k)\left(\varphi(\frac{h-k\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-k\rho}{\sqrt{1-\rho^2}})\right) - \rho z(m)\left(\varphi(\frac{h-m\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-m\rho}{\sqrt{1-\rho^2}})\right),$$

$$R \cdot E(y) = z(k)\left(\varphi(\frac{h-k\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-k\rho}{\sqrt{1-\rho^2}})\right) - z(m)\left(\varphi(\frac{h-m\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-m\rho}{\sqrt{1-\rho^2}})\right)$$
$$+ \rho z(h)\left(\varphi(\frac{k-h\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-h\rho}{\sqrt{1-\rho^2}})\right) - \rho z(l)\left(\varphi(\frac{k-l\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-l\rho}{\sqrt{1-\rho^2}})\right),$$

$$R \cdot E(x^2) = R + hz(h)\left(\varphi(\frac{k-h\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-h\rho}{\sqrt{1-\rho^2}})\right)$$
$$- lz(l)\left(\varphi(\frac{k-l\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-l\rho}{\sqrt{1-\rho^2}})\right)$$
$$+ \frac{\rho\sqrt{1-\rho^2}}{\sqrt{2\pi}}\left(z(\sqrt{\frac{h^2-2hk\rho+k^2}{1-\rho^2}}) - z(\sqrt{\frac{l^2-2lk\rho+k^2}{1-\rho^2}})\right)$$
$$- \frac{\rho\sqrt{1-\rho^2}}{\sqrt{2\pi}}\left(z(\sqrt{\frac{h^2-2hm\rho+m^2}{1-\rho^2}}) - z(\sqrt{\frac{l^2-2lm\rho+m^2}{1-\rho^2}})\right)$$
$$+ \rho^2 kz(k)\left(\varphi(\frac{h-k\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-k\rho}{\sqrt{1-\rho^2}})\right)$$
$$- \rho^2 mz(m)\left(\varphi(\frac{h-m\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-m\rho}{\sqrt{1-\rho^2}})\right),$$

$$R \cdot E(y^2) = R + kz(k)\left(\varphi(\frac{h-m\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-k\rho}{\sqrt{1-\rho^2}})\right)$$
$$- mz(m)\left(\varphi(\frac{h-m\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{l-k\rho}{\sqrt{1-\rho^2}})\right)$$
$$+ \frac{\rho\sqrt{1-\rho^2}}{\sqrt{2\pi}}\left(z(\sqrt{\frac{h^2-2hk\rho+k^2}{1-\rho^2}}) - z(\sqrt{\frac{m^2-2mh\rho+h^2}{1-\rho^2}})\right)$$
$$- \frac{\rho\sqrt{1-\rho^2}}{\sqrt{2\pi}}\left(z(\sqrt{\frac{l^2-2kl\rho+m^2}{1-\rho^2}}) - z(\sqrt{\frac{l^2-2ml\rho+m^2}{1-\rho^2}})\right)$$
$$+ \rho^2 hz(h)\left(\varphi(\frac{k-h\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-h\rho}{\sqrt{1-\rho^2}})\right)$$
$$- \rho^2 lz(l)\left(\varphi(\frac{k-l\rho}{\sqrt{1-\rho^2}}) - \varphi(\frac{m-l\rho}{\sqrt{1-\rho^2}})\right).$$

$$R \cdot E(xy) = \rho R + h\rho z(h)\left(\varphi\left(\frac{k-h\rho}{\sqrt{1-\rho^2}}\right)-\varphi\left(\frac{m-h\rho}{\sqrt{1-\rho^2}}\right)\right)$$

$$-l\rho z(l)\left(\varphi\left(\frac{k-l\rho}{\sqrt{1-\rho^2}}\right)-\varphi\left(\frac{m-l\rho}{\sqrt{1-\rho^2}}\right)\right)$$

$$+\frac{\sqrt{1-\rho^2}}{\sqrt{2\pi}}\left(z\left(\sqrt{\frac{h^2-2hk\rho+k^2}{1-\rho^2}}\right)-z\left(\sqrt{\frac{l^2-2ml\rho+m^2}{1-\rho^2}}\right)\right)$$

$$-\frac{\rho\sqrt{1-\rho^2}}{\sqrt{2\pi}}\left(z\left(\sqrt{\frac{l^2-2kl\rho+k^2}{1-\rho^2}}\right)-z\left(\sqrt{\frac{h^2-2hm\rho+m^2}{1-\rho^2}}\right)\right)$$

$$+\rho kz(k)\left(\varphi\left(\frac{h-k\rho}{\sqrt{1-\rho^2}}\right)-\varphi\left(\frac{l-k\rho}{\sqrt{1-\rho^2}}\right)\right)-\rho mz(m)\left(\varphi\left(\frac{h-m\rho}{\sqrt{1-\rho^2}}\right)-\varphi\left(\frac{l-m\rho}{\sqrt{1-\rho^2}}\right)\right).$$

One can invert these to calculate the parameters from the estimated sample moments. However, there is no explicit form for these "inverse" moment functions. Thus, numerical methods must be used. Those familiar with the singly truncated bivariate normal distribution or bivariate distributions truncated on only one variable (used heavily in models of sample selection) will recognize the general form of these equations and can verify that the equations given here converge to the equations for these more familiar distributions as the relevant limits of integration go to infinity or negative infinity. One interesting feature of these moment equations is that the correlation between x and y can be nonzero even if ρ is zero.

APPENDIX B
Derivation of the Likelihood Function

The presentation of the derivation of the likelihood function used in EI is relatively intuitive, but rather involved. To present the function, I first present the likelihood function that could be used if β^{D} and β^{R} were directly observed in the data. Let

$$\boldsymbol{\mu} = \begin{bmatrix}\mu_{\mathrm{D}}\\ \mu_{\mathrm{R}}\end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix}\sigma_{\mathrm{D}}^2 & \sigma_{\mathrm{DR}}\\ \sigma_{\mathrm{DR}} & \sigma_{\mathrm{D}}^2\end{bmatrix}.$$

Also, following King (1997, appendix D) let

$$R(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_0^1\int_0^1 f_n(x,z|\boldsymbol{\mu}\boldsymbol{\Sigma})dxdz.$$

The distribution for the truncated bivariate normal bounded by the unit square is then given by

$$f(\beta^{\mathrm{D}},\beta^{\mathrm{R}}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{f_n(\beta^{\mathrm{D}},\beta^{\mathrm{R}}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{R(\boldsymbol{\mu},\boldsymbol{\Sigma})}. \quad (7)$$

The distribution of the βs conditional on the observed P and D is found by exploiting well-known properties of the bivariate normal distribution. Note that

$$\begin{bmatrix}\beta^{\mathrm{D}}\\ P\end{bmatrix} = \begin{bmatrix}1 & 0\\ D & (1-D)\end{bmatrix}\begin{bmatrix}\beta^{\mathrm{D}}\\ \beta^{\mathrm{R}}\end{bmatrix}.$$

letting

$$\mathbf{A} = \begin{bmatrix}1 & 0\\ D & (1-D)\end{bmatrix}$$

and using well-known properties of linear transformations of normally distributed variables,

$$(\beta^{\mathrm{B}},P) \sim \frac{f_n(\beta^{\mathrm{B}},P|\mathbf{A}\boldsymbol{\mu},\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')}{R(\boldsymbol{\mu},\boldsymbol{\Sigma})}.$$

We can now factor this bivariate normal distribution into a conditional and marginal distribution, that is, $f_n(\beta^{\mathrm{B}}, P|\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}') = f_n(\beta^{\mathrm{B}}|P)f_n(P)$. Note that both the conditional and marginal distribution are themselves univariate normal distributions.

We now integrate over the feasible values of β^{D} to find the total likelihood of observing a precinct with the given P and D. Again following King, define

$$S(D,P,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_{\max(0,\frac{P+D-1}{D})}^{\min(1,\frac{P}{D})} f_n(\beta_{\mathrm{D}}|\mathbf{\Pi},\mathbf{A}\boldsymbol{\mu},\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')d\beta^B.$$

Thus, the full probability function defining $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in terms of the observables (P and D) is

$$f(P|D,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{S(P,D,\lambda,\mathrm{P})f_n(P|\mu_P,\sigma_P)}{R(\boldsymbol{\mu},\boldsymbol{\Sigma})},$$

where $\mu_P = D\mu_{\mathrm{D}} + (1-D)\mu_{\mathrm{R}}$ and $\sigma_P^2 = D^2\sigma_D^2 + (1-D)^2\sigma_{\mathrm{R}}^2 + 2(1-D)D\sigma_{\mathrm{DR}}$ taking logs and summing over i observations we find

$$\ln L = \sum_i \ln\frac{S(P_i,D_i,\boldsymbol{\mu},\boldsymbol{\Sigma})}{R(\boldsymbol{\mu},\boldsymbol{\Sigma})} + \sum_i \ln f_n(P|\mu_{Pi},\sigma_{Pi}).$$

To maximize $\ln L$, solve the first-order conditions, for example,

$$\frac{\delta \ln L}{\delta\mu_D} = \sum_i \frac{\delta}{\delta\mu_D}\ln\frac{S(P_i,D_i,\mu,\Sigma)}{R(\mu,\Sigma)} + 2\sum_i\left(\frac{P_i-\mu_{Pi}}{\sigma_{Pi}}\right)D_i = 0.$$

Note that the orthogonality condition is somewhat different from the one used in the ordinary least squares and general least squares methods used above to estimate the truncated moments. In those cases (putting aside truncation), $\sum_i \varepsilon_i D_i$ or

$$\sum_i \frac{\varepsilon_i D_i}{\sigma_{Pi}}$$

is set equal to 0; here $\sum_i \varepsilon_i D$ is set equal to

$$\sum_i \frac{\delta}{\delta\mu_{\mathrm{D}}}\ln\left[\frac{S(P_i,D_i,\mu,\Sigma)}{R(\mu,\Sigma)}\right].$$

If there is little truncation,

$$\sum_i \frac{\delta}{\delta\mu_D}\ln\left[\frac{S(P_i,D_i,\mu,\Sigma)}{R(\mu,\Sigma)}\right]$$

will be close to 0, and the untruncated and truncated parameters will be nearly identical.

APPENDIX C
Derivation of the Precinct Quantities of Interest

Using matrix notation, write

$$\begin{bmatrix}\beta^{\mathrm{D}}\\ P\end{bmatrix} = \begin{bmatrix}1 & 0\\ D & (1-D)\end{bmatrix}\begin{bmatrix}\beta^{\mathrm{D}}\\ \beta^{\mathrm{R}}\end{bmatrix}.$$

Define

$$\mathbf{A} = \begin{bmatrix}1 & 0\\ D & (1-D)\end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix}\sigma_{\mathrm{D}}^2 & \sigma_{\mathrm{DR}}\\ \sigma_{\mathrm{DR}} & \sigma_{\mathrm{R}}^2\end{bmatrix}.$$

Then putting aside the issue of truncation, the joint distribution of β^D and P would be

$$(\beta^D, P) \sim f_{bvn}(\beta^B, P; A\mu, A'\Sigma A).$$

Applying the truncation to this distribution is not straightforward, but fortunately it is unnecessary. Because we are only interested in the conditional distribution $f(\beta^D|P)$, we need only consider the truncation of β^D and not the truncation of P. Because the conditional distributions of bivariate normal variables are univariate normal, the untruncated distribution of β^D given P will be

$$\beta^D \Big| P \sim f_n\left(\beta^D; \mu_D + \frac{\sigma_{DP}}{\sigma_P}(P - \mu^P), \sigma_D - \frac{\sigma_{DP}}{\sigma_P}\right),$$

where $\mu_P = \mu_D D + \mu_R(1 - D)$, $\sigma_P = \sqrt{V_{\{2,2\}}}$, $\sigma_{DP} = V_{\{2,1\}}$, and $V = A'\Sigma A$. Rewriting, the untruncated distribution of β^D given P as $f_n(\beta^D; \mu_{D|P}, \sigma_{D|P})$, applying the bounds on β^D, and using a well-known formula for the expectation of doubly truncated univariate normal distributions, the expectation of β^D given P is

$$E(\beta^D \Big| P) = \left(\frac{f_n(\beta_l^D; \mu_{D|P}, \sigma_{D|P}) - f_n(\beta_u^D; \mu_{D|P}, \sigma_{D|P})}{F_n(\beta_u^D; \mu_{D|P}, \sigma_{D|P}) - F_n(\beta_l^D; \mu_{D|P}, \sigma_{D|P})}\right)\sigma_{D|P} + \mu_{D|P},$$

where $F_n(\cdot)$ is the standard normal cumulative distribution function and $\beta_l{}^D$ and $\beta^D{}_U$ are the lower and upper bounds on β^D. One can then make the estimate of β^R by applying equation (4).

Equation (8) reveals the effect of estimating each precinct at its expectation. If $\mu_{D|P}$ lies within the logical bounds, it will be the mode. From equation (8), we see that unless the bounds are symmetric about $\mu_{D|P}$ [i.e., f_n(lower bound β_i^D) = f_n(upper bound β_i^D)], the mode and mean will differ. If the truncation is greater on the high side (β^D large), the mean will be less than the mode, whereas if the truncation is greater on the low side, the mean will be greater than the mode. Because the estimate of β^R is decreasing in the estimate of β^D, the degree of polarization [abs($\beta^D - \beta^R$)] will always be smaller if the expectation (mean) and not the mode is used as the predicted value. This moderating effect may be in part responsible for EI's robustness to violations to its assumptions in empirical applications because such violations often lead to overestimates of polarization. One can make estimates of the Bs by applying equations (1) and (2) to the point estimates of the βs.

NOTES

1. In most applications, we cannot pare down the data to a 2 × 2 table without making some (possibly strong) assumptions. Take the example of making ecological inferences about race and vote choice. To fit these data into a 2 × 2 table, we must assume that the proportion of whites and nonwhites among those voting for third-party candidates or abstaining is the same as the proportion of whites and nonwhites in the total population.

2. Note that "Democrat" refers to individuals who vote for the Democratic assembly candidate.

3. A nice treatment of this problem is given in Ansolabehere and Rivers (n.d.). A more general though more technical treatment is given in Cho (1998).

4. As mentioned in more detail later, by assuming a parametric distribution for the βs, King (1997) is able to allow some dependence between the βs and D; however, this extension of his basic model depends much more heavily on the validity of this distributional assumption.

5. Of course, other assumptions could also buy us a lot, such as Freedman et al.'s (1991) contextual effects model in which it is assumed that $P_i = \beta_i^D = \beta_i^R$. This model completely determines the joint distribution of βs using only the marginal distribution of P!

6. Beran and Hall (1992) give a similar result for a somewhat more general problem. Given that Rivers has shown that the joint distribution of the βs is uniquely determined by the joint distribution of P and D under the assumption of independence, one might wonder why anyone would pursue King's parametric model. The short answer is efficiency. Because we generally have relatively few observations on P and D, making accurate nonparametric estimates of the joint distribution of the βs is not usually possible.

7. It should be noted that King does offer an extension of his model that relaxes this assumption. Some consideration of this extension is given below.

8. King reinforces this inference by claiming that his basic model "is robust even to high levels of aggregation bias" (p. 37). This claim is at the center of some recent controversy (see Freedman, Klein, and Ostland 1998 and subsequent correspondence).

9. The extensions of King's method that attempt to relax the independence assumption are beyond the scope of this study and are not considered in the applications given below.

10. Strictly speaking, King's estimator is fully Bayesian and not maximum likelihood.

11. Consistent estimates of B^D and B^R can be had from the ecological regression data without the assumption of independence of N and the βs if the regression weights the precincts by their size (number of voters), as shown by Ansolabehere and Rivers (n.d.).

12. Goodman (1959) considers such an estimator.

13. Although infeasible estimates of E^D and E^R may well indicate the presence of aggregation bias, the contrapositive is not true. Admissible estimates of E^D and E^R cannot be taken as evidence that there is no aggregation bias.

14. Of course, in real applications, it is exactly these quantities that we are trying to infer.

15. The parameters of the untruncated bivariate normal distribution are the means, variances (standard deviations), and covariance.

16. This is likely the result of Bayesian priors that King's (1997) ecological procedure places on the parameters.

17. Technically, King's (1997) method provides Bayesian rather than maximum-likelihood estimators for the TBVN parameters. King places prior distributions over each of the parameters. However, as generally implemented, these priors are weak (have very little effect on the estimates) and thus, without too much injustice, I present his method in terms of the more familiar one of maximum likelihood.

18. To simplify notation, I have suppressed the subscripts on β^D and β^R. This omission is innocuous in this case because the (β^D, β^R) pairs are assumed to be independently and identically distributed across precincts.

19. Because the precinct estimates are not consistent, we would not expect them to fall exactly on the 45° line, even as the size of sample grows large.

20. There are only 306, rather than 325, data sets because there are 19 pairs for which King's (1997) estimator failed to produce estimates. The likelihood function for King's method is difficult to maximize. For this reason, the computer procedure that implements his technique occasionally fails to produce estimates particularly for ill-conditioned data sets.

21. Although access to incredibly powerful computers has largely obviated the need for computationally efficient approximations, there are still some cases in which one might desire an estimator such as fastEI. For example, Lewis (1997) presents the results of over two thousand separate ecological inference problems. In this case, fastEI outpaces King's estimator by a matter of days.

22. In substantive applications, fastEI is a tool one might use for a quick and easy preview of what King's (1997) EI is likely to produce. However, because fastEI is inconsistent, users should employ King's estimator to produce final results.

23. By the so-called accounting identity—given in equation (4)—the error in the Republican support estimate will be $D/(1 - D)$ times as large as the error in the estimated Democratic support, where D is the share of voters who are Democrats districtwide. This factor holds exactly for the EI-type estimators and approximately for Goodman's ecological regression.

24. The cap on the maximum absolute error will be a function of how much information is contained in the bounds (see King 1997). For some sets of marginal distributions of P and D, the cap will be very close to the true values. For others, the cap could be very large.

25. This is well known in the literature (cf. King 1997; Achen and Shively 1995).

26. The maximum-entropy estimator of Johnston and Hay (1982) does allow for estimates of the individual precinct βs.

REFERENCES

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

Ames, B. 1994. The reverse coattails effect: Local party organization in the 1989 Brazilian presidential election. *American Political Science Review* 88(1):95–111.

Ansolabehere, S., and D. Rivers. N.d. Bias in ecological regression. Stanford University working paper.
Beran, R., and P. Hall. 1992. Estimating coefficient distributions in random coefficient regressions. *Annals of Statistics* 20(4):1970–84.
Cho, W.-T. 1998. Iff the assumptions fit. *Political Analysis* 7: 143–63.
Duncan, O.-D., and B. Davis. 1959. An alternative to ecological correlation. *American Journal of Sociology* 18:665-66.
Fréchet, M. 1951. Sur les tableaux de correlation dont les marges sont donnees. *Annals of the University of Lyon* 14:53–77. Section A, Series 3.
Freedman, D. A., S. P. Klein, and D. Ostland. 1998. Review of A solution to the problem of ecological inference. *Journal of the American Statistical Association* 93:1518–22.
Freedman, D. A., S. P. Klein, J. Sacks, C. A. Smyth, and C. G. Everett. 1991. Ecological regression and voting rights. *Evaluation Review* 15(6):672–711.
Goodman, L. 1959. Some alternatives to ecological correlation. *American Journal of Sociology* 64:610–25.
Griffiths, W. E. 1972. Estimation of actual response coefficients in the Hildreth-Houck random coefficient model. *Journal of the American Statistical Association* 67:633–35.
Hildreth, C., and J. P. Houck. 1968. Some estimators for a linear model with random coefficients. *American Statistical Association Journal* 63:584–95.
Johnston, R. J., and A. M. Hay. 1982. On the parameters of unified swing in single-member constituency electoral systems. *Environment and Planning* 14(1):61–74.
Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee. 1985. *The theory and practice of econometrics*. New York: Wiley.
King, G. 1997. *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton: Princeton University Press.
———. 1999. The future of ecological inference research. *Journal of the American Statistical Association* 94:352–55.
King, G., B. Palmquist, G. Adams, M. Altman, K. Benoit, C. Gay, J.-B. Lewis, R. Mayer, and E. Reinhard. 1997. The record of American democracy, 1984-1990. Cambridge, Mass.: Harvard University [producer]; Ann Arbor, Mich.: ICPSR [distributor].
Lewis, J. B. 1998. Who do representatives represent? The importance of electoral coalition preferences in California. Ph.D. diss., MIT.
Rivers, D. 1998a. Nonparametric estimation of ecological regression models. Paper presented at the American Political Science Association Meetings, Boston.
———. 1998b. Review of A solution to the problem of ecological inference. *American Political Science Review* 92(2):442–43.
Robinson, W. S. 1950. Ecological correlation and the behavior of individuals. *American Sociological Review* 15:351–57.
Schwallie, D. P. 1982. Unconstrained maximum likelihood estimation of contemporaneous covariances. *Economic Letters* 9:359–64.
Shah, S. M., and N. T. Parikh. 1964. Moments of the doubly truncated standard bivariate normal distribution. *Vidya* (Journal of Gujarat University) 7:81–91.