

The Statistical Foundations of the “EI” Method

Kenneth F. McCue*

California Institute of Technology

November 17, 1999

*Research Scientist, Environmental Quality Laboratory, California Institute of Technology, 138-78 Caltech, Pasadena, CA, 91125, 626-395-6869 (office), 626-395-2940 (fax), mccue@caltech.edu (e-mail).

Abstract

A recently proposed statistical model of ecological inference (the inference of individual behavior from aggregated data), known by its adherents as “EI”, has recently gained a great deal of attention both inside and outside the statistical profession. This article shows that “EI” is in fact an application of the standard statistical theory of prediction, though with many statistical errors, not the least of which is the failure of the author of “EI” to recognize the relationship between “EI” and prediction. While application of the theory of prediction may improve the case-level inference of individual behavior, it is unlikely to improve the estimates for the overall dataset, which is the usual goal of ecological inference.

1 Introduction

The field of ecological inference is concerned with the problem of determining the behavior of individuals with certain characteristics when this behavior is observed not at the level of the individual, but rather after the application of some aggregation mechanism. A simple form of the problem (which is sufficient for the following exposition) is that there exists a dataset of p electoral units, an electoral unit (usually referred to as a precinct) being the aggregation where the election authorities report totals of individual ballots. Now, each electoral unit i has associated with it N_i^b individuals with characteristic b , N_i^w individuals with characteristic w , and also the total number of votes for a particular candidate (call it Y_i). The ecological inference problem is then to determine what proportion of the voters with characteristic b voted for the candidate and what proportion with characteristic w voted for the candidate.

One solution is through some form of linear regression, estimating an equation of the form $Y_i = N_i^b b^b + N_i^w b^w + v_i$ or some variant thereof, with b^b being interpreted as the proportion of voters with characteristic b supporting the candidate and b^w being interpreted as the proportion of voters with characteristic w supporting the candidate (this was first done by Goodman (1953)). Another solution technique is known as the method of bounds (Duncan and Davis (1953)), which notes that for any individual precinct, b^b and b^w can only take on certain values, the set of which are often a much smaller interval than the $[0, 1]$ interval in which b^b and b^w , being proportions, can logically fall. There are difficulties with both approaches; linear regression estimates can fall outside of $[0, 1]$, while with the method of bounds, it is not obvious how to combine the individual ranges in order to obtain an overall estimate. A model which purports to combine these two methods in a manner which solves the problems of both has recently been introduced by King (1997), and is now known in the social science statistical literature as “EI.”

The use of a acronym to described a statistical technique of course has strong connotations to statisticians, as powerful techniques throughout statistics are generally known by their abbreviations (EM, MLE, and OLS come to mind). It is therefore not unnatural for statisticians to wonder whether

knowledge of “EI” is something that a statistician needs to add to his or her professional repertoire. Certainly the publicity surrounding this statistical technique might lead one to think so. For example, the flyer issued by Princeton University Press (1997) described the publication of King’s book as being “the subject of unprecedented press releases from the National Science Foundation and the Council of Scientific Society Presidents,” and describes the book as “a solution to the longest standing methodological problem in political science research.” Furthermore, accounts of his proposed methodology have even appeared in the popular press, such as the *New York Times* (1997).

In spite of these press releases and topical converge, however, it is probably safe to say that most statisticians were unaware of the existence of “EI” until an unusual exchange in *The Journal of the American Statistical Association* between King (1999) and Freedman, Ostland, Roberts, and Klein (1999). This exchange was occasioned by a review of King (1997) by Freedman, Klein, Ostland and Roberts (1998), the essence of the review being that, in the time-honored tradition of ecological inference, it was possible to find datasets for which the “EI” procedure gave poor estimates of ecological behavior without being able to determine, through the “EI” diagnostic procedures, that these results were lacking. Furthermore, the review made extensive comparisons to a model known as the neighborhood model (originally proposed by Freedman et. al (1991), and which is not a statistical model at all but rather an allocation rule), and concluded that the neighborhood model produced superior results to “EI”. There was very little discussion of the “EI” procedure as a statistical procedure (essentially one paragraph in (Freedman et. al. (1998), p. 1521), and no discussion on the claims of King’s to have discovered what King considered to be a new and powerful form of statistical inference (see King (1997, pp. 91-92), for example).

As Freedman et. al. (1998) could not properly be considered a review of King (1997), the book review editor at *JASA* invited King to prepare a response to the review and this appeared in the “Letters to the Editor” section (King (1999)). While this was a rare opportunity for King to present the essence of his model in its most appealing form to the statistical profession (and what author would not avail himself of the the opportunity to write a review of his own book?), his reply did not do that; rather, it primarily

consisted of two parts, the first of which was a discussion of the motives of Freedman et. al (1991, 1998) for advancing the neighborhood model, the second of which was a series of simulated examples as to how Freedman et. al. (1998) had misapplied the diagnostics present in the “EI” procedure. Simulation was necessary due to the refusal of Freedman et. al (1998) to make their data available; Freedman et. al. (1999) responded that the original refusal to share data had come from King, and Freedman et. al. (1999) provided their own evidence that they had indeed applied the diagnostics correctly. The cycle of example, counter example and counter counter example was now complete.

While the exchange between Freedman et. al. (1998, 1999) and King (1999) provided an interesting insight into the dynamics behind the usually dry process of academic discourse (a senior editor at *JASA* I spoke with used the term “personality”), the problem with press releases, polemic exchanges and mutual accusations is that they tend to obscure the statistical aspects of any procedure. And while there has been a great deal of obscuration on the “EI” method, the essence of the King “EI” method is quite simple and already exists in the statistical literature, for, fundamentally, the King “EI” method is a straightforward application of the statistical theory of prediction. This article, then, briefly outlines the King method in relation to that theory, discusses some of King’s claims of improved efficiency and robustness for district-wide estimators, and then discusses some of the factors which have led, up to this point, to the failure to place the King “EI” procedure in its proper place in statistical theory.

2 The King “EI” method

The basics of the King model is summarized by Wang (1998) and that summary is followed here (in King (1997) the method is spread out over multiple chapters and appendices and is quite difficult to follow). Essentially, there are two types of individuals in an electoral precinct i , type b (for black) and type w (for white). There are N_i^b blacks and N_i^w whites in the precinct, for N_i voters overall. Useful “district-wide” quantities (a district is a collection of precincts) are $N^b = \sum_{i=1}^p N_i^b$ and $N^w = \sum_{i=1}^p N_i^w$, and $N = N^b + N^w$,

where p is the number of precincts in the district. Turnout (the act of casting a ballot, regardless of candidate choice) in the precinct can be expressed as $N_i^T = N_i^b \beta_i^b + N_i^w \beta_i^w$, or $T_i = X_i \beta_i^b + (1 - X_i) \beta_i^w$ (the accounting identity), where $T_i = N_i^T / N_i$ and $X_i = N_i^b / N_i$, whence $1 - X_i = N_i^w / N_i$ (King actually develops a complicated method of dealing with both turnout and vote choice through separate equations and then combining the results from each equation, but the “EI” method can be illuminated by only dealing with the turnout equation.) If there are p precincts, there are p pairs (β_i^b, β_i^w) , or $2p$ unknowns. Without additional assumptions no unique solution exists, as an infinite set of pairs of (β_i^b, β_i^w) solve the set of equations.

King’s solution is to assume that the (β_i^b, β_i^w) are distributed bivariate truncated normal on the unit square and to estimate the parameters of the underlying bivariate normal distribution (call them $\check{\psi} = (\check{\mathcal{B}}^b, \check{\mathcal{B}}^w, \check{\sigma}_b^2, \check{\sigma}_w^2, \check{\sigma}_{bw})$) by the method of maximum likelihood. Suitable assumptions are made to ensure that the usual product likelihood function is valid, and for the moment it will be assumed that the King model is the correct specification. King adopts a two-stage procedure, first estimating parameters of the underlying bivariate normal and then using these estimates to calculate precinct-level estimates, which he then combines to create district-level estimates. The complexity of this procedure demands extensive numerical integration be performed (King uses simulation), but as will be seen later, a procedure exists that achieves much the same results as the King procedure, with this alternative procedure using only numeric optimization of the likelihood and the usual application of asymptotic limit theorems. The existence of this type of procedure, being primarily analytic, can then be used to examine King’s claims as to the robustness and increased efficiency of the “EI” methodology.

This reliance of the precinct-level estimators on the maximum likelihood estimators can be seen in King’s bifurcated approach to estimation, as he first estimates the parameters of the bivariate normal by the method of maximum likelihood and then (usually, see King (1997, sec. 7.4 and 7.5)) uses the sampling distributions of the mles as a prior for a second stage of estimation, where he estimates

$$\Pr [\beta_i^b | T] = \int \Pr [\phi(\check{\psi}) | T] \Pr [\beta_i^b | T, \check{\psi}] d\check{\psi}, \quad (1)$$

the ϕ being a deterministic transformation of the $\check{\psi}$ chosen for its numeric

tractability (King, 1997, p. 136). Since ψ is a five-dimensional vector, this is a complicated integral and a numeric evaluation must be performed (this evaluation is described in King (1997, sec. 8.2) and is involved). His precinct-level estimators are thus based on the precinct-level random variables $\beta_i^b|T_i$, which function almost like primitives in the King model. Bounds on the support of $\beta_i^b|T_i$ can be derived by taking extreme values of β_i^w in the accounting identity ($T_i = X_i\beta_i^b + (1 - X_i)\beta_i^w$), and these bounds correspond to the usual method of bounds in the ecological literature (letting β_i^w equal zero or one gives $\beta_i^b \in [\text{Max}(0, T_i/X_i - (1 - X_i)/X_i), \text{Min}(1, T_i/X_i)]$, for example). Thus King can plausibly claim that his method of estimation combines the method of Goodman (linear regression, or King’s maximum likelihood model here), with the method of bounds. Respects is probably a better word than combines, since the assumption that the accounting identify holds, plus restricting (β_i^b, β_i^w) to be bivariate truncated normal on the unit square, ensures that the Duncan and Davis bounds will obtain and both of these assumptions are built into the likelihood.

Since King believes $\beta_i^b|T_i$ is the best predictor function of β_i^b when making precinct-level inferences (“By far the best summary of our knowledge of the parameter [King means β_i^b] ... is the entire posterior distribution [i.e., $\beta_i^b|T_i$] (King, 1997, p. 149)),” it follows that graphs of this distribution provide the most information, and they are duly provided at places in his work (see King, 1997, p. 148, for example). This belief in the optimality of $\beta_i^b|T_i$ is at odds with statistical theory, however, which holds that the best predictor function for β_i^b would be $E(\beta_i^b|T_i)$, not $\beta_i^b|T_i$ (in terms of mean square error—see a text such as Bickel and Doksum (1977, sec. 1.6), for example). King’s justification for his preference for $\beta_i^b|T_i$ is given in King (1997, sec. 6.2.3), and this justification bears some resemblance to the usual statistical derivation of the best predictor function of one variable by another, insofar as King does (for the case of the bivariate normal) compare variation between $\text{Var}(\beta_i^b|T_i)$ and $\text{Var}(\beta_i^b)$, and finds that variance of the first quantity is less than that of the second. The use of the bivariate normal as an example has apparently mislead King, though, since for that distribution the quantity $\text{Var}(\beta_i^b|T_i)$ is not a function of T_i , something which is not usually the case. The random quantity present in $\text{Var}(\beta_i^b|T_i)$ for most distributions would make categorical statements such as that cited at the beginning of this paragraph equivocal at best; most comparisons between random variables rely on some type of

averaging.

When it comes to actually calculating precinct-level point estimates, King himself uses the averaging in the expectation operator to create these estimates. King choice is the usual estimator from statistical theory, $E(\beta_i^b|T_i)$, and he simulates the expectation of the probability distribution displayed in (1) by taking multiple draws from the distribution and averaging these together (he suggests a possible use of the median but appears to use the mean almost exclusively—the median, of course, would put his estimates away from the optimal predictor function). While this procedure might be expected to produce a consistent estimator of the expected value of $\beta_i^b|T_i$ (assuming $\beta_i^b|T_i$ is calculated correctly, a point that will be discussed later), King also uses these draws to compute the standard deviation based on the deviation of these draws from the average of these draws. This is usually incorrect, as the desired variation is that of $E(\beta_i^b|T_i)$, not the variation of $\beta_i^b|T_i$, which is what King’s procedure provides an estimate of.

3 Calculation of Precinct-level Estimators

For actually calculating the precinct-level estimators, first the case of the bivariate normal will be considered, then the truncated bivariate normal. For the bivariate normal, of course, no simulation is necessary for the calculation of the precinct-level estimators, as all of these quantities can be calculated analytically. If (β_i^b, β_i^w) are bivariate normal, $E[\beta_i^b|T_i]$ is simply $\check{\beta}^b + \check{\alpha}_i^b[T_i - \check{\mu}_i]$, where $\check{\mu}_i = X_i\check{\beta}^b + (1 - X_i)\check{\beta}^w$ and $\check{\alpha}_i^b = \text{Cov}[\beta_i^b, T_i|\check{\psi}]/\text{Var}[T_i|\check{\psi}]$ (see Morrison, section 3.4, for a complete derivation under the condition of multivariate normality). Under normality, then, the $\hat{\beta}_i^b$ is a positive affine transformation of $T_i - \check{\mu}_i$, so that $\hat{\beta}_i^b$ is a linear function of the residual. Thus the information in any of King’s $\hat{\beta}_i^b$ is no different than the information in the residual for the case when (β_i^b, β_i^w) is bivariate normal, and any of King’s diagnostic procedures which rely upon the estimated $\hat{\beta}_i^b$ are in reality relying upon the residual of the regression, which will make them no more or no less effective than a diagnostic procedure which relies upon the residual.

Turning to the case where (β_i^b, β_i^w) is truncated bivariate normal, let $\psi =$

$(\mathcal{B}^b, \mathcal{B}^w, \sigma_b^2, \sigma_w^2, \sigma_{bw})$ be the parameters of that distribution. Then $\mathcal{B}^b + \alpha_i^b [T_i - \mu_i]$, where $\mu_i = X_i \mathcal{B}^b + (1 - X_i) \mathcal{B}^w$ and $\alpha_i^b = \text{Cov} [\beta_i^b, T_i | \psi] / \text{Var} [T_i | \psi]$, is still a good estimator for $\hat{\beta}_i^b$, in that $\beta_i^b - \hat{\beta}_i^b$ continues to have zero expectation and the above definition of α_i^b minimizes the variation of this random difference for linear functions of the T_i . It can also be shown that this estimator satisfies the accounting identity (using $X_i \text{Cov} (\beta_i^b, T_i) + (1 - X_i) \text{Cov} (\beta_i^w, T_i) = \text{Var} [T_i]$ and some algebraic manipulation) and that $\hat{\beta}_i^b$ and $\hat{\beta}_i^w$ are both in the unit interval (by taking the extreme points of (T_i, X_i) and mapping them into the $(\hat{\beta}_i^b, \hat{\beta}_i^w)$ plane). So for precinct-level estimates that use the linear form of the estimator, the bounds provided by Duncan and Davis are satisfied. This linear estimator is a function of the parameters of the truncated bivariate normal; estimated parameters from a numeric optimization of the likelihood would be those of the underlying (untruncated) bivariate normal. To convert between these two sets of parameters, one must use analytical formulas such as those given in Regier and Hamdan (1971) (King, incidentally, is under the misapprehension that the precise functional relationship between the parameters of the underlying bivariate normal distribution and the truncated bivariate normal distribution cannot be solved analytically (King (1997), p. 105)). The formulas given in this paper allow the usual application of asymptotic theory to find confidence intervals for the linear estimator when the parameters are replaced with the maximum likelihood estimates.

Using $E [\beta_i^b | T_i]$ is of course better in terms of minimal variation but the complexity of calculating this quantity often argues against its use. Such is the case with the truncated bivariate normal, as King miscalculates the distribution of $\beta_i^b | T_i$ (which he then uses to simulate $E [\beta_i^b | T_i]$). King claims that $\beta_i^b | T_i$ is truncated univariate normal with the mean of the underlying normal distribution being $\check{\mathcal{B}}^b + \check{\alpha}_i^b [T_i - \check{\mu}_i]$ and the variance being $\text{Var} [\beta_i^b] - \check{\alpha}_i^b \text{Cov} [\beta_i^b, T_i | \psi]$. To obtain this result, King (1997, app. C) treats (β_i^b, β_i^w) as bivariate normal, then transforms (β_i^b, β_i^w) to (β_i^b, T_i) , which is also bivariate normal. The mean and variance of the (normal) conditional variate $\beta_i^b | T_i$ is then calculated in the usual way. At this point, however, King truncates $\beta_i^b | T_i$ (he refers to as “putting off truncation until the end” (King 1997, p. 306)) to obtain the truncated univariate normal distribution. These operations are invalid if (β_i^b, β_i^w) are bivariate truncated normal, as truncation in general cannot be “put off” or altered. As an example, the sum of two normal variables truncated after summation is truncated normal but the sum of two

truncated normal variables is not truncated normal, a fact which can be established through manipulation of characteristic functions.

That this distributional derivation is invalid can also be seen from the fact that the equality $E [E [\beta_i^b | T_i]] = E [\beta_i^b]$ does not hold, since by equation (79) in Johnson and Kotz (1970), $E [\beta_i^b | T_i] = \check{\mathcal{B}}^b + \check{\alpha}_i^b (T_i - \check{\mu}_i) + \epsilon(T_i)$, the ϵ being a term which depends upon the limits of integration. Then $E [E [\beta_i^b | T_i]] = \check{\mathcal{B}}^b + \check{\alpha}_i^b [(\mathcal{B}^b - \check{\mathcal{B}}^b) X_i + (\mathcal{B}^w - \check{\mathcal{B}}^w)(1 - X_i)] + E [\epsilon(T_i)]$, which is not, in general, equal to \mathcal{B}^b unless (β_i^b, β_i^w) are distributed bivariate normal, whence $E [\epsilon(T_i)] = 0$, $\check{\mathcal{B}}^b = \mathcal{B}^b$ and $\check{\mathcal{B}}^w = \mathcal{B}^w$. Thus all of the substantive conclusions King draws from his interpretation of $\beta_i^b | T_i$ as being truncated univariate normal distribution (and there are a number, see King (1997, p. 108), for example) are wrong. Use of the correct $E [\beta_i^b | T_i]$ would require simulating $\beta_i^b | T_i$ and then simulating the expectation of this variate; it is not clear that after all of this simulation one would have a better estimator than the linear one derived above and even less clear that there is any substantive difference, as $E [\beta_i^b | T_i]$ will almost certainly be close to the linear estimator, except, perhaps, in extreme cases. This implies that the $\hat{\beta}_i^b$ based on this expectation will be close to being a linear function of the residual, so even if $E [\beta_i^b | T_i]$ were calculated correctly the $\hat{\beta}_i^b$ derived from using it would still be basically a function of the residual.

4 Calculation of District-level Estimators

District-wide estimates of individual behavior is the usual goal of ecological inference, and King claims that the use of precinct-level estimators improves the estimation of district-wide parameters, specifically mentioning that the preferred estimator for the percent of blacks voting is $\sum N_i^b \beta_i^b / N^b$ (King (1997), p. 32). Since this is a random function and not a statistic, King takes $\hat{\beta}_i^b$ and replaces β_i^b in this random function to create a statistic. Using the linear precinct-level estimator derived above, this statistic can be written as

$$\frac{\sum N_i^b \hat{\beta}_i^b}{\sum N_i^b} = \mathcal{B}^w + \sum \frac{N_i^b}{N^b} \alpha_i (T_i - \mu_i)$$

(recall $N^b = \sum N_i^b$). Asymptotically, if the N_i^b/N^b are all negligible, the law of large numbers causes the last term to go to zero and King's district-wide estimator reduces to \mathcal{B}^w , which is best estimated by use of the appropriate maximum likelihood estimator, not King's estimator. King, however, justifies the use of this estimator on the basis of the existence of one or more N_i^b/N^b being non-negligible (the existence of such terms has implications for the use of maximum likelihood estimation, which also depends on the law of large numbers, though King does not discuss this). If it is decided to approximate $\sum N_i^b \beta_i^b / N^b$, King's substitution of $\hat{\beta}_i^b$ for β_i^b is inefficient. To see this, let $\theta_i = N_i^b / N^b$, $r_i = \theta_i^2 \alpha_i^2 \text{Var} [T_i - \mu_i]$, and $s_i = \theta_i \alpha_i \text{Cov} [\beta_i^b, T_i - \mu_i]$. For the best linear unbiased estimator of $\sum \theta_i \beta_i^b$ (which is King's estimator), the method of LaGrange multipliers can be used to find γ_i such that the random function $\sum \theta_i \beta_i^b - \sum \gamma_i \hat{\beta}_i^b$ has zero expectation and minimal variation (the solution is $\gamma_i = (2s_i + \lambda) / 2r_i$, where $\lambda = (\sum 1/2r_i)^{-1} [1 - \sum (s_i/r_i)]$).

Suppose now it is assumed that King's model does not represent the data correctly, that is, the "EI" underlying probability model is misspecified. King makes claims as to the robustness of his procedure, and the analytic framework for King's procedure established above offers a way to evaluate this for district-wide quantities. Under a bivariate normal distribution (without truncation), consider the case where the true model is $T_i = X_i \beta_i^b + (1 - X_i) \beta_i^w = X_i \beta_i = X \beta + u_i$, $\text{Var} (u_i) = \sigma_i^2$, but the model estimated is $T_i = Z_i \tau_i = Z_i \tau + v_i$, $\text{Var} (v_i) = \lambda_i^2$. Assuming a random sample of size n , let X be the n by 2 stacked matrix of the X_i 's, Z be the n by G stacked matrix of the Z_i , Y_i be the n by 1 stacked matrix of the Y_i , $\beta = (\tilde{\beta}^1, \tilde{\beta}^2)$ be a 2 by 1 mean vector of the bivariate normal, and τ be a G by 1 vector, τ being of course a function of β . Letting Λ being the n by n matrix with the variance terms located on the diagonal, off-diagonal entries being zero, then the log likelihood of this specification is

$$\begin{aligned}
\ln L [Z|\tau] &= c - \sum \ln \lambda_i + .5(Y - Z\tau)' \Lambda^{-1} (Y - Z\tau) \\
&= c - \sum \ln \lambda_i + .5(Y - X\beta)' \Lambda^{-1} (Y - X\beta) \\
&\quad + (Y - X\beta)' \Lambda^{-1} (X\beta - Z\tau) \\
&\quad + .5(X\beta - Z\tau)' \Lambda^{-1} (X\beta - Z\tau)
\end{aligned} \tag{2}$$

(conditions for convergence of this type of misspecified likelihood are given in such texts as White (1994)). Differentiating the last expression with respect

to τ and equating to zero, one obtains

$$0 = Z'\Lambda^{-1}(Y - X\beta) + Z'\Lambda^{-1}(X\beta - Z\tau), \quad (3)$$

or

$$\begin{aligned} \hat{\tau} &= (Z'\Lambda^{-1}Z)^{-1}Z'\Lambda^{-1}X\beta \\ &+ (Z'\Lambda^{-1}Z)^{-1}Z'\Lambda^{-1}(Y - X\beta) \end{aligned} \quad (4)$$

Scaling $(Z'\Lambda^{-1}Z)^{-1}$ by n and $Z'\Lambda^{-1}X\beta$ and $Z'\Lambda^{-1}(Y - X\beta)$ by $1/n$ and taking limits gives

$$\text{plim } \hat{\tau} = \left[\text{plim } \frac{(Z'\Lambda^{-1}Z)}{n} \right]^{-1} \text{plim } \frac{(Z'\Lambda^{-1}X)}{n} \beta,$$

as the last term of (4) consists of deviations from zero assumed to be uncorrelated with any other variable. This estimator is the same as the GLS estimator under the incorrect specification (the variance parameters will need to be estimated also but the above will hold for those estimated values).

To examine the robustness of this estimator, let $M_i^g = Z_i^g N_i$ be the number of people with characteristic g in precinct i , τ_i^g be the random coefficient for the i^{th} precinct, $\hat{\tau}_i^g$ be the estimator of this random coefficient, and $\hat{\tau}^g$ be the g^{th} component of $\hat{\tau}$. Then using the linear form of the precinct-level estimator, since $T_i - \mu_i = X_i\beta + u_i - Z_i\tau$, the district-wide quantities for any g can be expressed as

$$\begin{aligned} \frac{\sum M_i^g \hat{\tau}_i^g}{\sum M_i^g} &= \hat{\tau}^g + (\sum M_i^g)^{-1} \\ &\sum M_i^g \frac{\text{Cov} [\tau_i^g, T_i]}{\text{Var} [T_i]} [X_i\beta + u_i - Z_i\tau]. \end{aligned} \quad (5)$$

To show robustness under a violation of the model assumptions, one would need to show that this expression is “closer” to the true value than the one under a regression. To set a situation where this judgment can be made, consider the case where both X_i and Z_i have two groups, say latino and non-latino, but there has been errors in the classification, with some latinos being treated as non-latinos and vica versa (if the classification of latino

ethnicity is done by surname matching, as it often is, some error is inevitable, if only due to intermarriage and some women taking the surname of their husband). Then robustness here would be that the King district estimator, $\sum M_i^1 \hat{\tau}_i^1 / \sum M_i^1$ is closer to $\beta^1 = \check{\beta}^1$ from the true model than it is when $\hat{\tau}^1$ is obtained from a regression procedure. Inspection of (5) gives no particular reason to believe that this estimator would be robust in that sense, and there is a case where the King district-wide estimator and the GLS estimator can be shown to be the same. Let (X_i, Z_i) be independently, identically distributed multivariate normal variates. Then from (3) it follows that from construction of the τ in the maximization, the quantity $(X_i\beta - Z_i\tau)/\lambda_i^2$ is normal with mean zero and has zero covariance with Z_i , and is hence independent (by normality) of any function of Z_i . $\text{Cov}[\tau_i^g, T_i]$, is, however, only a function of the Z_i , as it is calculated under the assumption that $Y_i = Z_i\tau_i$ is the true model (the randomness in T_i is removed via the expectation operator). So the second term of (5) will approach zero, and in this case the King estimator reduces to a generalized least squares estimator, and the King procedure is not robust at the district level (giving the same answer as generalized least squares). The robustness of King’s district-level estimator to other forms of misspecification and/or violations of assumptions can also be analyzed using the framework described here.

5 The “EI” Method and Statistical Theory

From the above, it is clear that King’s methodology is not a new finding in statistical theory, and essentially his methodology reduces to allocating the residual from his linear regression type procedure to the individual coefficients of his varying parameters regression. While these precinct-level estimators (when correctly calculated) will be better for the prediction of precinct results than simply using the mean from the underlying distribution (and will also share diagnostic properties of the residuals of the regression), inference on the underlying parameters of the distribution will not be improved and there is no reason to expect improvements in the district-wide quantities, either in terms of estimation or robustness under model misspecification. The “EI” model can thus be seen as a straight-forward application of the mathematical theory of prediction, even though not perceived as such

by the author.

This misperception comes about for a number of reasons. First, there is a lack of statistical knowledge and sophistication in King's work. Aside from the various probabilistic errors outlined above (which can be corrected), and an apparent lack of knowledge of a common statistical procedure, King simply assumes and/or asserts too often. As one example, he assumes that because his estimators work better at the precinct level than the estimated mean form the truncated bivariate normal, any district estimator constructed from these estimates must also be more accurate at the district level, though he calculates no covariance matrices to show that this is the case. Or as another example, he believes that there will be a fundamental improvement in his estimation of $\beta_i^b|T$ by treating his estimates of $\check{\psi}$ as uncertain and conditioning on them, as in (1). This approach can be compared to the usual method of asymptotic inference, which substitutes the mle estimator of $\check{\psi}$ into the distribution function of $\beta_i^b|T, \check{\psi}$, and uses this distribution for inference. When King complains (King, 1999, p. 353) "They [Freedman et. al. (1998)] would condition on uncertain point estimates of intermediate parameters, rather than correctly include the full uncertainty of all quantities," he is claiming that his method is superior to standard asymptotic inference. It may be, but he provides no support for this assertion. A priori, it seems unlikely that there would be any real difference between estimates from the two approaches, particularly since King is already accepting asymptotic methods by his use of the method of maximum likelihood. On the other hand, there would be tremendous simplification in his model if the usual asymptotic approximation were used.

The second reason for King's failure to identify his procedure in the statistical literature is the ease with which simulation is now done, given the recent advances in software and hardware. Statistics by simulation is the wave of the future, and it seems inevitable that it will replace the often-difficult probabilistic manipulations which historically have served as the basis for statistical theory (King notably states that "Simulation ... is a lot easier, faster, and more intuitive than the corresponding mathematical derivations (1997, p. 141)"). While there is certainly a place for every technique, it seems all too likely that a reliance on simulation distorts the practitioner's view of statistical theory. As an example, consider the following (King (1997,

p. 107)):

For example, most textbook presentations of regression analysis first derive the variance of the coefficient vector conditional on the regression variance. In the usual notation, $V(b|\sigma^2) = (X'X)^{-1}\sigma^2$. Only after this calculation do we average over the uncertainty in σ^2 to yield the unconditional variance $V(b) = (X'X)^{-1}\hat{\sigma}^2 = (X'X)^{-1}e'e/(n - k)$.

Needless to say, textbook presentation of simple linear regression do not consider the variance parameters of the regression σ^2 to be uncertain—rather, it is assumed to be a fixed parameter which happens to be unknown. Would King conceptualize linear regression in this way if he were not treating other fixed parameters as random to make his calculations? How many of the other mistakes that he makes might have been caught if analytic means were the primary method of analysis, rather than simulation, or even if certain analytical checks, such as calculating expectations, were carried out?

Finally, any theory does not exist in a vacuum—it has adherents and detractors. Ideally, in a scientific environment, one's degree of adherence or dissension is related to scientific principles. The degree of “hype” on this theory, however, leads one to believe that at some points this is not the case. King (1999, p. 352) himself speaks of the “excitement” which his theory has generated, and there is now a cottage industry applying it (King (1999) gives a number of references); these researchers have a vested interest in having “EI” accepted as the primary methodology of ecological inference. The role of statisticians in cases like this should be to make certain that sound statistical theory is applied, and this article is an effort to make certain that is the case. Since King's (1997) work passed through peer review at a major press and has been the subject of many presentations and academic papers in the field of political methodology, all without either the errors discussed in this article or the statistical basis of “EI” being uncovered, it is clear that it is going to take statisticians to provide statistical guidance. After all, the statistical technique of prediction, when carried out correctly, would certainly be useful to improve precinct-level results. The district-wide claims for “EI”, however, seem to be unsupported by statistical theory.

Articles Referenced.

- Advertising Flyer, Princeton University Press (1997). For *A Solution to the Ecological Inference Problem*.
- Bickel, Peter J. and Doksum, Kjell A.(1977). *Mathematical Statistics*. San Francisco, CA: Holden-Day, Inc.
- Duncan, O.D. and Davis, B. (1953) “An Alternative to Ecological Correlation”, *American Sociological Review*, 18: 665-666.
- Goodman, L. (1953). “Ecological Regressions and the Behavior of Individuals”, *American Sociological Review*, 18: 663-664.
- Freedman, D. A., Klein, S. P., Sacks, J., Smyth, C. A., and Everett, C. G., (1991). “Ecological Regression and Voting Rights,” *Evaluation Review*, 15, 673-711.
- Freedman, D. A., Klein, S. P., Ostland M. and Roberts, M. R. (1998). Review of *A Solution to the Ecological Inference Problem*, *Journal of the American Statistical Association*, 93(444), December, 1518-1522
- Freedman, D. A., Ostland, M. A, Roberts, M. R. and Klein, S. P. (1999). “Response to King”, Letter to the Editor in *Journal of the American Statistical Association*, 94(445), March, 355-357
- Freeman, Karen (1997). “Statistician Builds What May Be a Better Data Mousetrap”, *The New York Times*, July 8, 1997: page C8
- King, Gary. (1997). *A Solution to the Ecological Inference Problem*. Princeton, New Jersey: Princeton University Press
- King, Gary. (1999). “The Future of Ecological Inference Research: A Comment on Freedman et. al.”, Letter to the Editor in *Journal of the American Statistical Association*, 94(445), March, 352-355.
- Johnson, Norman L. and Kotz, Samuel. (1970). *Continuous Univariate Distributions - 1*, Boston: Houghton Mifflin Company.
- Morison, (1976) *Multivariate Statistical Methods*. McGraw-Hill Book Company, New York.

- Regier, Mary and M.A. Hamdan. (1971). ‘Correlation in a Bivariate Normal Distribution with Truncation in Both Variables,’ *Australian Journal of Statistics*, 13(2), 77-82.
- Wang, Fang. (1997). “An Alternative Conjecture–Guide to King’s”, *The Political Methodologist*, 8(1), 21-26.
- White, (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press. Cambridge, UK.