THE FUTURE OF ECOLOGICAL INFERENCE RESEARCH: A COMMENT ON FREEDMAN ET AL.

I appreciate the editor's invitation to reply to the review by Freedman, Klein, Ostland, and Roberts (1998) of *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior From Aggregate Data* (King 1997). I welcome this scholarly critique and *JASA*'s decision to publish in this field. Ecological inference is a large and very important area for applications that is especially rich with open statistical questions. I hope that this discussion stimulates much new scholarship. Freedman et al. (1998) raise several interesting issues, but also misrepresent or misunderstand the prior literature, my approach, and their own empirical analyses, and have compounded the problem by refusing requests from me and the editor to make their data and software available for this note. Some clarification is thus in order.

I believe that the excitement surrounding the publication of my book resulted from its unifying in a single statistical approach the insights of two literatures that had been in open warfare since the 1950s. Unlike Freedman et al.'s (1998) implication, I never claimed that all problems were resolved in my book (which is why the title referred to "*a* solution," rather than "*the* solution"). Numerous important issues remain, and much research is needed. With an area as uncertain and important as ecological inference, making available many solutions, formalized as statistical models, is critical for applied researchers.

Freedman et al. (1998) are right about one point: if one can avoid making inferences about individuals from aggregate data, then one should do so. And of course, valid survey data make ecological inferences superfluous. Unfortunately, Freedman et al. (1998) do not consider the many researchers who *must* make ecological inferences, even when risky. This position is consistent with the low value that Freedman generally seems to put on model-based inference, such as in regression analysis for causal inference (Freedman 1998) or sampling adjustments for the U.S. Census (Brown et al. 1998), and may explain why Freedman et al. sometimes appear ambivalent about making ecological inferences even with their own model (Freedman, Klein, Sacks, Smyth, and Everett 1991, pp. 682, 806). As statisticians, if they feel uncomfortable with the assumptions, they can work on other problems or conclude that "surveys offer a better approach" (p. 701), but many applied researchers have no such options. For example, implementing the U.S. Voting Rights Act, and thus court decisions about how to hold democratic elections, requires ecological inferences, because surveys on race are known to be misleading. Surveys are not available for most historical questions, such as who voted for the Nazi Party in Weimar Germany. Ecological inferences are required for preliminary studies in fields like epidemiology for use in deciding which expensive individual-level data to collect. They are also essential for such fields as marketing, education, sociology, political science, and others due to confidentiality or data limitations. Even when limited survey data are available, ecological or combined inferences are necessary for studying small geographic areas.

Freedman et al.'s (1998) review is primarily concerned with comparing the method proposed by Freedman et al. (1991) to that in my book (now known in the literature as EI) and Goodman's (1953, 1959) regression. Predictably, Freedman et al. (1991) think that their own approach outperforms both competitors. Because some information is lost to aggregation, it is no surprise that they were able to find examples for which one of the EI models gives inaccurate answers. Indeed, my book includes a 35-page section (titled "What Can Go Wrong?") with far more egregious examples that they seem to have missed. The question is not whether one can find data for which a statistical model will perform poorly, as this would be easy in almost any area of statistics. The questions instead are how much information a model uses, how often one can recognize when the model is inappropriate, and whether the approach suggests other models in these instances. I show here that EI offers progress in each area.

This is Freedman et al.'s (1998) third attempt at promoting their "neighborhood model," either in its own right or as a foil to stop others from making ecological inferences. They introduced it in expert witness testimony, but it was dismissed out of hand ["As such, it is not a reliable method of inferring group voting behavior," (Garza et al. v. County of Los Angeles, 1990, p. 35)], primarily because its key assumption, described later, is both unreasonable and untestable. It was offered in the social sciences but was rejected as "politically naive" (Grofman 1991) and painting "a picture of America that no one would recognize—because it does not exist" (Lichtman 1991).

In 1991, both Grofman and Lichtman discovered that Freedman et al. (1991) presented a biased sample of estimates and datasets that supported their approach and suppressed the remaining evidence that was contradictory. Freedman et al. did the same in 1998 by reporting estimates of (minority) Hispanic behavior but suppressing estimates for non-Hispanic behavior, about which there is typically more information of the type EI might have extracted. However, their selective presentation cannot be corrected or completed without access to their data. Although we cannot know how often EI diagnostics could have detected problems with either their reported or unreported estimates, I show here with simulated data that match their requirements how EI can help in similar situations.

From 1953 to 1997, the methods of Goodman (1953) and Duncan and Davis (1953) were the only practical, useful, and thus widely used methods, despite an active methodological literature. In 1991 Freedman et al. proposed the neighborhood model as part of an attack on Goodman. But to my knowledge, this model has never been used in applied research, and applications of Goodman's regression did not decline. In contrast, because it uses more information, allows more palatable assumptions, offers a range of specifications, and combines existing approaches, EI has seen use since 1997 in a number of academic disciplines and in industry and public policy.

BACKGROUND

Following the running example in my book, the ecological inference problem uses the fraction of persons of voting age who are black (X_i) and who vote (T_i; and the total number, N_i) to estimate the fraction of blacks (β_i^b) and whites (β_i^w) who vote in precinct i. A precinct is a voting booth and its associated geography; p precincts compose a legislative district. The secret ballot prevents computing the cross-tabulation directly. Because precincts have known identities (unlike, say, anonymous survey respondents), real applications come with considerable qualitative information (largely ignored by Freedman et al. 1991).

Since 1953, two hotly contested approaches have dominated the literature: those following the monumental contributions of Goodman (1953, 1959) and Duncan and Davis (1953). Goodman showed that the accounting identity $T_i = X_i\beta_i^b + (1 - X_i)\beta_i^w$ holds and gave the assumptions necessary to generate valid estimates of the district-wide fractions of blacks ($B^b = \sum_{i=1}^p N_i X_i \beta_i^b / \sum_{i=1}^p N_i X_i$) and whites [$B^w = \sum_{i=1}^p N_i(1 - X_i)\beta_i^w / \sum_{i=1}^p N_i(1 - X_i)$] who vote by a regression of T_i on X_i and $(1 - X_i)$ with no constant. Its key assumption is $\text{cov}(\beta_i^b, X_i) = \text{cov}(\beta_i^w, X_i) = 0$. Freedman et al.'s caricature of Goodman's assumption as "constancy" of β_i^b and β_i^w over i is thus wrong (and easily rejected by the presence of any scatter around a regression line). As Goodman made clear, his approach reveals no information about whether

Gary King is Professor, Department of Government, Harvard University, Cambridge, MA 02138 (E-mail: *King@Harvard.edu*). The software for ecological inference mentioned in this letter is available at *http://GKing.Harvard.Edu*.

Journal of the American Statistical Association
March 1999, Vol. 94, No. 445, Letters to the Editor

this assumption holds. For example, letting the coefficients B^b and B^w each be linear functions of X_i produces a quadratic regression (of T_i on a constant, X_i, and X_i^2), but the resulting three coefficients do not uniquely identify the four necessary parameters (the intercepts and slopes for B^b and B^w). If its assumption is wrong, then estimates from Goodman's regression can be wrong and even outside [0,1].

Duncan and Davis's (1953) approach provides deterministic bounds on β_i^b and β_i^w narrower than $[0, 1]$: $\beta_i^b \in [\max(0, \{[T_i - (1 - X_i)]/X_i\}), \min(1, (T_i/X_i))]$ and similarly for β_i^w. How much information this provides depends on the application, but it can be substantial. For example, if precincts in a scatterplot of X_i by T_i are uniformly spread over the unit square, then the average width of the bounds on β_i^b is reduced from [0,1] to less than one-half that range, hence eliminating 50% of the ecological inference problem with certainty. The information also varies over i, so that we might have $\beta_1^b \in [.1, .4]$, $\beta_2^b = .8$, $\beta_3^b \in [.5, .6]$, $\beta_4^b \in [0, 1]$, and so on. Each line in what I called a tomography plot [because of the equivalence between ecological inference and certain medical imaging problems; see Fig. 1(a)] portrays this information in terms of all values of β_i^b and β_i^w consistent with X_i and T_i for precinct i. Ecological inference may have more uncertainty than other inferences, due to the information lost to aggregation, but it also offers remarkable certainty for some quantities of interest.

Freedman et al.'s (1991) neighborhood model assumes that $\beta_i^b = \beta_i^w$; that is, black and white turnouts are equal. This assumption narrows estimates of each to T_i with a 0 standard error. (Thus, applying their "Z score" evaluation to their own model, as they did to EI in table 1, indicates worse performance than all other approaches.) The model could be formulated as approximate with meaningful standard errors, but this has not been done. Researchers have found this model useless, as it assumes the answer to the question, for example, of whether black and white turnouts are equal.

Freedman et al.'s (1991) model is a clever way to restate the indeterminacy of the problem: Because T_i always falls within the bounds, the assumption cannot be rejected from the data alone. Precinct estimates are given by where the corresponding tomography line crosses the 45-degree diagonal. Among the many other possible untestable assumptions is Flanigan and Zingale's earlier (1985) proposal to use the midpoint of the bounds. The fully general result has been given by King (1997, p. 191): an assumption about the joint density of β_i^b and β_i^w cannot be rejected from the data as long as it has mass over any curve that connects the bottom left and top right points of a tomography plot. This fundamental indeterminacy is one way of expressing the information lost due to aggregation. (Chap. 3 gives the others.)

THE EI APPROACH

The key feature of EI is that it uses information from *statistical*, *deterministic*, *diagnostic*, and *qualitative* sources, whereas previous approaches use only one of the first two. Although supporters of Goodman (1953, 1959) and of Duncan and Davis (1953) have been at each other's throats for years, I showed that the two approaches can be combined in a way that improves on each—first extracting the (often substantial) unit-level deterministic information and then statistically borrowing strength to narrow in probabilistically within these bounds. This formalizes what the best applied analysts do informally by applying the unit-level bounds as post hoc checks on Goodman's regression (a key point that Freedman et al. missed in both 1991 and 1998). The basic EI model has three assumptions given X_i: (a) β_i^b and β_i^w are truncated bivariate normal on the unit square; (b) T_i and $T_{i'}$ are independent for all i not equal to i'; and (c) the dependence of β_i^b and β_i^w on X_i must be assumed or estimated in some way. The accounting identity and these assumptions imply a likelihood for T_i. Conditioning on T_i after estimation yields a joint posterior for β_i^b and β_i^w on its tomography line, which can be used to compute posteriors for B^b and B^w, if desired. [Freedman et al.'s (1998) "fiducial twist" comment misrepresents how these are computed and misunderstands how they should be computed; they would condition on uncertain point estimates of intermediate parameters, rather than correctly including the full uncertainty of all quantities.]

Including the Duncan–Davis bounds in a statistical model makes EI more efficient than Goodman's method. That it can also make EI more robust can be shown in several ways, all ignored by Freedman et al. (1998) but essential in real applications.

1. My Monte Carlo experiments show that many violations of (a) are unimportant for B^b and B^w and that violations of (b) are innocuous even for fairly high levels of spatial correlation. Freedman et al.'s (1998) claim that EI cannot recover the right parameter values from data simulated from EI's model is wrong. The book includes examples to the contrary; I and others have run numerous other tests; and EI code is free, open, and on the Web. Clearly, Freedman et al.'s home-grown simulation routines are in error, but what mistake they have made is a mystery, because, along with their data, their code is also not available.

2. Suppose that in (c) X_i is assumed independent of β_i^b and β_i^w. Then as $\text{cov}(X_i, \beta_i^b)$ or $\text{cov}(X_i, \beta_i^w)$ increases from 0, the absolute bias of Goodman's regression increases linearly without limit, but for EI it increases linearly at the start, then slows, and finally reaches a maximum and stops (King 1997, p. 180). How fast the maximum is reached depends on how informative are the bounds—which is known from the aggregate data. Thus the risks are knowable ex ante if, unlike Freedman et al. (1998) one looks before running an analysis. This specification uncertainty should be added to the computed model uncertainty before drawing conclusions.

3. Because the bounds are used, β_i^b and β_i^w can be functions of X_i without Goodman's identification problem.

EI also offers several diagnostics. Freedman et al. (1998) are concerned that I provide no examples that fit their model but not the basic EI model, and where EI diagnostics reveal the problem with EI. Indeed, the book does not include such an example, but it is easy to provide one. Lacking their data, I simulated some that meet their requirements. In these data B^b, B^w is .23, .13, which is estimated by the neighborhood model as .23, .14 and by the basic EI model [with independence for (c)] as .26, .02 (with standard errors .01, .02). What can be detected without knowing the true values? Figure 1 provides three diagnostics; by projecting lines onto the corresponding axis, the tomography plot (a) shows that most precincts

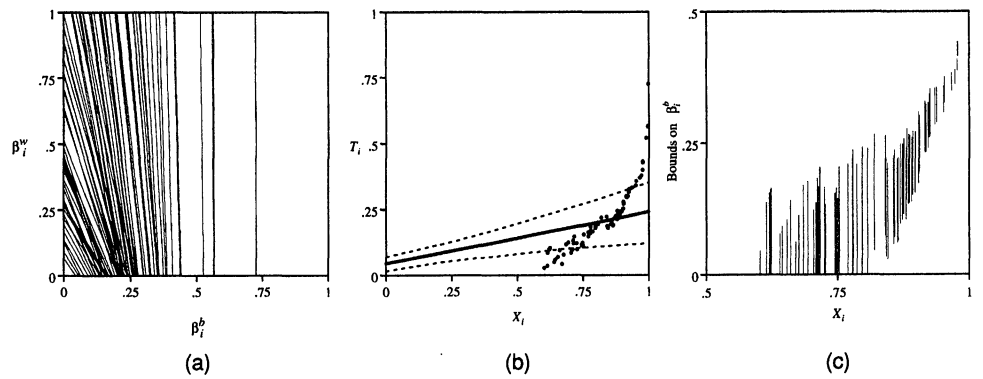

Figure 1. Some EI Diagnostics. (a) A tomography plot with lines $\beta_i^w = T_i/(1-X_i) - \beta_i^b X_i/(1-X_i)$ (b) $E(T_i|X_i)$ with 80% confidence intervals that do not fit the X_i, T_i points; and (c) a plot showing evidence that $\text{cov}(\beta_i^b, X_i) > 0$ from the bounds alone.

provide narrow bounds on β_i^b but wide bounds on most of the β_i^w's, with some (in the lower-left corner) providing narrow bounds on both. Measurement error aside, these bounds are assumption-free 100% confidence intervals (rare in any statistical field); although ignored by Freedman et al. (1998), precincts with sufficiently narrow bounds can be an immensely valuable resource for real problems. The district-wide bounds also reflect a similar pattern, with $B^b \in [.14, .26]$ and $B^w \in [0, .55]$, and so inferences about B^w must be more model-dependent than those about B^b. Figure 1(b) shows that $E(T_i|X_i)$ with 80% confidence intervals do not fit the X_i and T_i points, and so something is wrong. Figure 1(c) plots X_i by the (vertical) bounds on β_i^b, demonstrating that wherever the β_i^b's fall, $\text{cov}(\beta_i^b, X_i) > 0$, and so the basic EI model assumption of independence is rejected. Although no guarantee given the loss of information, a formal test (ignored by Freedman et al. 1998; see King 1997, p. 178) for whether X_i should be a covariate to control for this correlation is easily passed. (The coefficient on the covariate is more than five standard errors from 0.) This EI specification gives accurate estimates of .26, .13 with standard errors .02, .07, and nominal confidence interval coverage of β_i^b, β_i^w (for all i) at least as wide as the actual coverage.

Because EI keeps the posterior density within the bounds, Figure 1(c) shows how estimates of β_i^b and β_i^w can be correlated with X_i even if they are a priori independent, an important component of robustness and model checking. This plot also shows how adding the information from the Duncan–Davis (1953) bounds can greatly improve the Goodman (1953, 1959) approach, which yields estimates of .41, $-.61$ and is not even identified with X_i as a covariate.

The diagnostics will not always detect problems, but this example shows that they are helpful in some data. Whether they would help in Freedman et al.'s (1998) data is impossible to tell, because Freedman et al. fail to present most diagnostics, misinterpret warning messages that they generated by choosing incorrect specifications, use irrelevant tests like whether the regression of T_i on X_i is significant, and consider the effects of heteroscedasticity where it does not matter much (B^b and B^w) but ignore where it often does matter (β_i^b and β_i^w).

The final source of information that EI uses is qualitative, a critical source that Freedman et al. (1998) ignore. Because information is lost to aggregation, the only way to improve inferences after using deterministic information in X_i, T_i, and N_i is through external sources. Qualitative information may include ethnographies, participant observations, intensive interviews, partial survey data, journalistic accounts, historical studies, prior quantitative research, detailed information on a few precincts, and the like—the full range of data collection schemes used in the social sciences. If one is unsure of the answers of a survey respondent, then one is out of luck. If one is unsure of the results from the precinct or of an assumption in an ecological inference model, then qualitative information can be essential. At worst, one can go to the precinct and look around! The importance of qualitative information has been undervalued in the statistical literature on this problem, and perhaps even more generally except in some discussions of Bayesian analysis. However, it has been a long-standing concern among social science methodologists and applied researchers (e.g., King, Keohane, and Verba 1994). The EI approach includes this information as an integral part of the process of inference, and the software facilitates this process by allowing priors and many other specification options. This enables the researcher to add external information at numerous points in an analysis. Even if qualitative information led one to believe that $\beta_i^b = \beta_i^w$ was approximately right, using EI with this prior would better represent the available information than a degenerate posterior for each spiked at $T_i = \beta_i^b = \beta_i^w$, as Freedman et al. (1998) recommend.

Letting qualitative information guide model specification is a major deviation from Goodman's (1953, 1959) approach, which for all practical purposes offers only one option. Unlike the impression left by Freedman et al. (1998), EI results are not what come spinning out of the software with all 50+ options set at their defaults. If one is unsure of some assumptions, one should evaluate the sensitivity of conclusions to these options. Valid ecological inferences require considerable thought and evaluation of the many types of evidence that are available as part of real applications. Freedman et al. (1998, p. 1521) consider qualitative information only in their discussion of what covariates to use, but instead of doing the hard qualitative research themselves, they try only a sample of variables listed in one passage in my book (King 1997, p. 171). However, they neglect to mention that these variables were not uncovered by my qualitative research; indeed, I was merely quoting from their 1991 work, in which they offered qualitative justification for why unit-level parameters should vary over i as in their neighborhood model. Apparently, their 1998 analyses demonstrate that their 1991 justifications were wrong. Qualitative information is essential in ecological inferences, but it is valuable only if, as for quantitative data, substantial care and thought goes into its collection.

EI AND CHALLENGES FOR STATISTICAL RESEARCH

Goodman's regression works when its assumptions hold but not otherwise. No information exists within this framework to tell whether its assumptions are right or wrong. In contrast, under EI we now have five scenarios, only the last of which is a problem:

1. If the assumptions are correct, then one gets the right answer (e.g., King 1997, chaps. 9 and 10).
2. If the assumptions are wrong, then EI still does "well" (in the sense of mean squared error or absolute bias) when the bounds are sufficiently informative; the degree to which the bounds are informative can easily be assessed from the aggregate data, and so the risks are known ex ante (e.g., King 1997, chap. 11). Informative bounds improve inferences whether or not Goodman's model would have given estimates outside [0, 1].
3. If the assumptions are wrong, and the bounds are not sufficiently informative but the diagnostics are, then the assumptions can be changed and EI will do well (e.g., Fig. 1 herein and King 1997, pp. 176–180, 187, 195, 238).
4. If the basic assumptions are wrong and the bounds and diagnostics are not sufficiently informative, but the researcher has additional qualitative knowledge, then more appropriate assumptions can be chosen. Then either EI will do well, or the formal measures of uncertainty produced by EI (which are conditional on the model) can be supplemented and expanded accordingly.
5. If the assumptions are wrong and the bounds and diagnostics are not sufficiently informative, and the researcher has no time to collect qualitative information, then EI will perform poorly (e.g., King 1997, fig. 9.2). Even in this worst-case scenario, EI will be more robust than the Goodman approach, because the maximum bias from EI is fixed and knowable.

The odds of Scenarios 1–4 occurring relative to 5 (as compared to the odds of the assumptions applying versus not applying under Goodman's approach) summarize the advantages of EI. Thus EI chips pieces off of Goodman's worst case (the assumptions not applying). The benefits of EI are thus application dependent, but known ex ante.

I believe that EI is popular among researchers making ecological inferences because it uses more information—building on the fundamental approaches of both Goodman (1953, 1958) and Duncan–Davis (1953), including diagnostics that can often detect when assumptions are wrong, and adding a large range of models to the researcher's toolkit so that qualitative information can be included and sensitivity to assumptions can be assessed. In a word, more information is better. The neighborhood model is a useful reminder about indeterminacies and uncertainties in ecological inference, but it is untestable, and its assumption is unacceptable to applied researchers. Although methodologists have begun to build on the EI approach with hierarchical modeling (King, Rosen, and Tanner 1999), faster method of moment estimators (Lewis 1998), extensions to panel inference from independent cross-sectional surveys (Schuessler and Penubarti 1997), and nonparametric models (Rivers 1998), applied researchers in many fields could benefit from other developments. Some open issues include quantitative diagnostics, model selection, flexible distributional and functional form specifications, tractable extensions to multiple category and continuous individual-level variables, models for survey and ecological data, spatial autocorrelation, and the types of nonrandom measurement error common in ecological data. For statisticians willing to help the numerous researchers working on a diverse array of important problems with ecological data, many opportunities remain.

Gary KING

REFERENCES

Brown, L. D., Eaton, M. L., Freedman, D. A., Klein, S. A., Olshen, R. A., Wachter, K. W., Wells, M. T., and Ylvisaker, D. (1998), "Statistical Controversies in Census 2000," Technical Report 537, University of California, Berkeley.

Duncan, O. D., and Davis, B. (1953), "An Alternative to Ecological Correlation," *American Sociological Review*, 18, 665–666.

Flanigan, W. H., and Zingale, N. (1985), Alchemist's Gold; Inferring Individual Relationships From Aggregate Data, *Social Science History*, 9, 73–91.

Freedman, D. A. (1998), "From Association to Causation: Some Remarks on the History of Statistics," Technical Report 521, University of California Berkeley.

Freedman, D. A., Klein, S. P., Ostland, M., and Roberts, M. R. (1998) "Review of *A Solution to the Ecological Inference Problem*, by G. King," *Journal of the American Statistical Association*, 93, 1518–1522.

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.

Garza v. County of Los Angeles, 756 F. Supp. 1298 (U.S. Fed. Dist. Court, Calif. Central Dist., 1990).

Goodman, L. (1953), "Ecological Regressions and the Behavior of Individuals," *American Sociological Review*, 18, 663–666.

——— (1959), "Some Alternatives to Ecological Correlation," *American Journal of Sociology*, 64, 610–624.

Grofman, B. (1991), "Statistics Without Substance: A Critique of Freedman et al.," *Evaluation Review*, 15, 746–769.

King, G. (1997), *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior From Aggregate Data*, Princeton, NJ: Princeton University Press.

King, G., Keohane, R. O., and Verba, S. (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton, NJ: Princeton University Press.

King, G., Tanner, M. A., and Rosen, O. (1999), "Binomial-Beta Hierarchical Models for Ecological Inference," unpublished manuscript submitted to *Sociological Methods and Research*.

Lewis, J. (1998), "Method-of-Moment Estimators for King's Ecological Inference Model," paper presented at the Annual Meeting of the Social Science History Association.

Lichtman, A. J. (1991), "Passing the Test," *Evaluation Review*, 15, 770–799.

Rivers, D. (1998), "A Nonparametric Method for Ecological Inference," paper presented at the Annual Meeting of the American Political Science Association.

Schuessler, A., and Penubarti, M. (1997), "Individual Change From Independent Surveys via Ecological Inference," paper presented at the Annual Meeting of the American Political Science Association.

RESPONSE TO KING'S COMMENT

King (1999) has replied to our review of his book. After summarizing the issues, we will respond to the main points and a few of the minor ones. The book proposes a method for ecological inference (EI) and makes claims about its validity. According to King, his model provides realistic estimates of uncertainty, with diagnostics capable of detecting failures in assumptions. He also claims that the model is robust even when the assumptions are wrong.

In our review (Freedman, Klein, Ostland, and Roberts 1998) we showed by example that claims are seriously exaggerated. King's method works if its assumptions hold; if assumptions fail, estimates are unreliable, as are internally generated estimates of uncertainty. Diagnostics cannot distinguish between cases where his model works and where it fails.

MODEL COMPARISONS

Our review compared King's method to ecological regression and the neighborhood model. In our test data, the neighborhood model was the most accurate, whereas King's method was no better than the ecological regression. To implement King's method, we used his software package EZIDOS, which we downloaded from his web site. For a brief description of the EI and EZIDOS software packages, see King (1997 p. xix).

King (1999, p. 352) contends that we (1) used a biased sample of datasets and (2) suppressed "estimates for non-Hispanic behavior, about which there is typically more information of the type EI [King's method] would have extracted." Grofman (1991) and Lichtman (1991) are cited to support claim (1). Our answer is simple: we used the data that we had. Of course, Grofman and Lichtman made other arguments, too; our response is in Freedman et al. (1991).

We turn to point (2). It is not clear what sort of additional information would be available to King for non-Hispanics. Moreover, the neighborhood model and King's method get totals right for each geographical unit; thus any error on the Hispanic side must be balanced by an error of the same size but the opposite sign on the non-Hispanic side. (It is errors in the counts that balance; for ecological regression with unit weights, the balance is only approximate.) In short, King's method is unlikely to do better on non-Hispanics than it does on Hispanics.

Empirical proof will be found in Tables 1 and 2, which show results on non-Hispanics for the real datasets considered in our review. (Artificial data will be discussed later.) These tables and similar ones in our review show King's method to be inferior to the neighborhood model, for non-Hispanics as well as for Hispanics. Surprisingly, in the Los Angeles data, the results in Table 1 show that his method is inferior to ecological regression.

King (1997) tried his model on five datasets. These are not readily available, but we were able to get one of them—poverty status by sex in South Carolina block groups—directly from the U.S. Census Bureau. We ran the three ecological inference procedures on this dataset (Tables 1 and 2). King's method succeeds only in the sense that the estimate is within 1.1 standard errors of truth; the neighborhood model comes much closer to the mark, both for men and women. Where comparisons are feasible, the neighborhood model was more accurate than King's method on the real datasets, even in his own South Carolina example.

King (1999, p. 352) states that the neighborhood model is not a reliable method of inferring the behavior of subgroups from aggregate data; it is unreasonable, politically naive, and paints "a picture of America that no one would recognize." We would make two points in response: (1) the neighborhood model demonstrates that ecological inferences are driven largely by assumptions and not by data, a point that King almost concedes (1999, p. 354), and (2) the neighborhood model outperforms the competition, including King's method. If King's method does not perform as well as the neighborhood model, then his model cannot be described as reasonable, reliable, or politically savvy.

Diagnostics

King contends (1999, p. 352) that we (1) "misinterpret warning messages ... generated by choosing incorrect specifications," and (2) "use irrelevant tests like whether the regression of T_i on X_i is significant ...." (In the South Carolina example, T_i would be the fraction of persons in block group i who are below the poverty line, and X_i would be the fraction of persons in that block group who are male.) Both points simply misread what we wrote. With respect to (1), we interpreted the warning messages as evidence of specification error. With respect to (2), consider for instance Figure 1(b) in our review. The vertical axis shows $\hat{p}_i$ not T_i, an estimated propensity for a group rather than an observed fraction. This figure is one of King's "bias plots" (King 1997, p. 183). The issue is the regression of $\hat{p}_i$ on X_i, not the regression of T_i on X_i. The bottom line is that King's diagnostics raise warning flags even when his standard errors are reasonable, as in the Stockton results; equally, diagnostics are passed when the method fails, as in the Los Angeles results.

We now consider diagnostics for King's South Carolina data. Figure 1, (a) and (b), plots for each block group the estimated fractions of men and women in poverty against the fraction of men. (Every tenth block group is shown; estimates are computed by using King's software package EZIDOS.) The regression line for men has a shallow but statistically significant slope; the line for women falls quite steeply. King's assumption of independent and identically distributed propensities is strongly rejected by the data. Likewise, the warning messages point to specification error:

```
Warning: Some bounds are very far from distri-
bution mean. Forcing 2163 simulations to their
closest bound.
```