A rejoinder to King

2 March 1999

by

D. A. Freedman, Statistics Department, U. C. Berkeley, CA 94720

S. P. Klein, RAND Corporation, 1700 Main Street, Santa Monica, CA 90401

M. Ostland, Statistics Department, U. C. Berkeley, CA 94720

M. R. Roberts, Economics Department, U. C. Berkeley, CA 94720

Introduction

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 and makes sweeping 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 the model is robust even when assumptions are wrong.

We showed, by example, that these claims are seriously exaggerated. King's method works if its assumptions hold. If assumptions fail, estimates are unreliable: so 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, while King's method was no better than 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) contends that we (i) used a biased sample of data sets and (ii) 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 (i). 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 (ii). It is by no means 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 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 data sets 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 Hispanics. Surprisingly, in the Los Angeles data, his method is also inferior to ecological regression.

Table 1. Comparison of Three Methods for Making Ecological Inferences, in Situations where the Truth is Known; Results for non-Hispanics in Stockton and Los Angeles, and Men and Women in South Carolina

	Neighborhood model	Ecological regression	King's method	Truth	Z
Stockton					
Exit Poll	39.8	25.8	36.5 ± 3.6	42.0	−1.5
Los Angeles					
Education	76.4	81.6	82.9 ± 0.2	78.1	24.0
High Hispanic	60.1	71.9	73.1 ± 1.0	66.3	6.7
Income	53.5	55.4	56.4 ± 0.2	53.2	14.2
Ownership	56.1	57.4	57.5 ± 0.3	56.4	3.9
Party affiliation	58.6	57.2	54.6 ± 0.1	57.3	−33.0
High Hispanic	68.1	54.5	53.5 ± 0.4	61.5	−18.2
South Carolina					
Men in poverty	15.0	−13.3	5.8 ± 6.6	12.9	−1.1
Women in poverty	15.7	43.7	24.2 ± 6.1	17.7	1.1

NOTE: Values in percentages. King's method gives an estimate and a standard error, reported in the format "estimate ± SE." Z = (estimate − truth)/SE, computed before rounding. In South Carolina, block groups with fewer than 25 inhabitants are excluded from the data.

King (1997) tried his model on five data sets. 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 Census Bureau. We ran the three ecological-inference procedures on this data set (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 has been more accurate than King's method on the real data sets, even in his own South Carolina example.

King says 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: (i) the neighborhood model demonstrates that ecological inferences are driven largely by assumptions not by data—a point that King almost concedes; and (ii) the neighborhood model outperforms the competition, including King's method. If King's method performs even worse than ours, surely his model cannot be described as reasonable, reliable, or politically savvy.

Table 2. Which Estimation Procedure Comes Closer to Truth?

	Group	
	Hispanics	Non-Hispanics
Stockton		
Exit Poll	Nbd	Nbd
Los Angeles		
Education	Nbd	Nbd
High Hispanic	Nbd	Ecoreg
Income	Nbd	Nbd
Ownership	Nbd	Nbd
Party affiliation	Nbd	Ecoreg
High Hispanic	Nbd	Nbd
	Males	Females
South Carolina		
Poverty	Nbd	Nbd

NOTE: "Nbd" is the neighborhood model and "Ecoreg" is ecological regression. King's method does not appear in the table because in each case it does less well than the neighborhood model; furthermore, in each of the Los Angeles data sets, it does less well than ecological regression.

Diagnostics

King contends that we (i) "misinterpret warning messages ... generated by choosing incorrect specifications," and (ii) "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 (i), of course we interpreted the warning messages as evidence of specification error. With respect to (ii), 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 plot" (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: King's diagnostics raise warning flags even when his standard errors are reasonable, as in Stockton; equally, diagnostics are passed when the method fails, as in Los Angeles.

We now consider diagnostics for King's South Carolina data. Figures 1(a) and 1(b) plot 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 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 IID propensities is strongly rejected by the data. Likewise, the warning messages point to specification error:

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

King (1997, p. 225) insists that "even in [the South Carolina] data set, chosen for its difficulty in making ecological inferences, the inferences are accurate." But warning signals from the diagnostics have been ignored. Perhaps the idea is just this: when his method succeeds, it succeeds despite the difficulties; when it fails, it fails because of the difficulties.

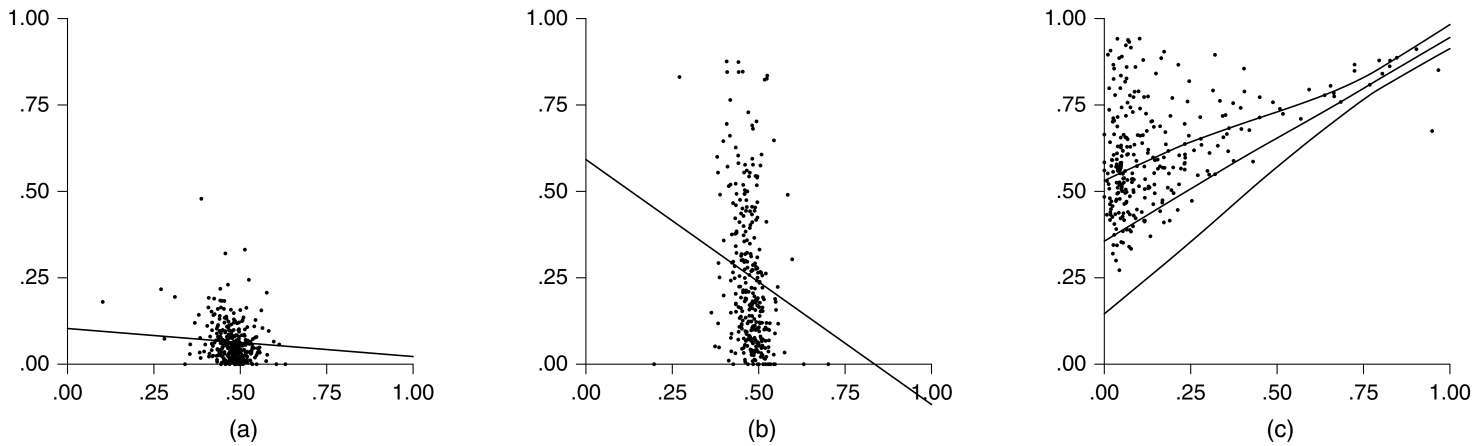

Figure 1. Diagnostic plots. (a) Bias plot for men in poverty, South Carolina. The plot shows $(x_i, \hat{p}_i)$. There is one dot per block group; the fraction x_i of people in block group i who are male is on the horizontal axis and the estimated fraction $\hat{p}_i$ of men in block group i who are poor is on the vertical axis. The slope of the regression line is small but significant. (b) Bias plot for women in poverty, South Carolina. The plot shows $(x_i, \hat{q}_i)$. There is one dot per block group; the fraction x_i of people in block group i who are male is on the horizontal axis and the estimated fraction $\hat{q}_i$ of women in block group i who are poor is on the vertical axis. The slope of the regression line is large and significant. (c) $E\{t|x\}$ plot showing (x_i, t_i); artificial data, Los Angeles. There is one dot per tract. The fraction x_i of people in tract i who are hispanic is on the horizontal axis and t_i, the fraction of people in tract i who register as democrats is on the vertical axis. Also shown are 80% confidence bands derived from the model; the middle line is the estimated $E\{t|x\}$. The dots are much too high, indicating a coding error in EZIDOS. Every tenth block group is shown for the South Carolina data, and every fifth tract for Los Angeles.

King imputes to us the "claim that EI cannot recover the right parameter values from data simulated from EI's model.". That is also a misreading. Of course King's method should work if its assumptions are satisfied—as we said on p. 1518 of our review, and demonstrated with two artificial data sets (pp. 1519–20). We still think there is a bug in King's software, because the diagnostics sometimes indicate problems where none can exist (p. 1520). Here is an example. Applied to the Los Angeles data on party affiliation, King's method estimates the five parameters of the untruncated normal distribution (two means, two SDs, and r) as 1.0456, 0.2853, 0.1606, 0.3028, −0.9640. We generated pairs of propensities from this bivariate distribution, kept only pairs that fell into the unit square, computed corresponding tract-level observations, and fed the resulting data back into EZIDOS. The parameter estimates were fine—1.0672, 0.2559, 0.1607, 0.3024, −0.9640.

The trouble comes in the diagnostics. Figure 1(c) shows our simulated data for every fifth tract. The figure also shows the 80%-confidence bands for the tract-level observations (the fraction who register democratic); the middle line is the conditional mean, estimated from our data—along with the bands—by EZIDOS. Clearly, something is wrong. The midline should more or less cut through the middle of the scatter diagram, and the band should cover about 80% of the dots. However, most of the dots are above the midline: indeed, about half of them spill over the top of the band. Similar errors are discussed by McCue (1998).

King presents artificial data for which his diagnostics pick up a failure in assumptions. This is an existence proof: there are some data sets for which the diagnostics work. Interestingly enough, the data had to be generated for the purpose. In the examples we considered, both real and artificial, the diagnostics were not reliable guides to the performance of King's method. Figure 1 above reinforces this point, for one of his own data sets (South Carolina), and for artificial data generated from his model (Los Angeles).

Other issues

King emphasizes throughout his reply that qualitative information needs to be used, the "50+ options" in his code being tuned accordingly. (Some options in EZIDOS allow for Bayesian inference rather than likelihood methods; others change the numerical algorithms that will be used; still others control print formats.) However, it is hard to see how qualitative information plays any role in the real examples presented by King (1997); and we saw nothing there about the 50+ options. On the contrary, the discussion of the real examples suggests straight-ahead use of maximum likelihood estimation.

King contends that our description of the constancy assumption is a "caricature." However, equation (2) in our review is exactly the one that is estimated by proponents of ecological regression, like Grofman and Lichtman. Moreover, King appears to misread Goodman (1953), who delineates the narrow circumstances under which ecological inference may be expected to succeed. We can all agree that, coldly stated, the assumptions underlying ecological regression are unbelievable.

King denies any "fiducial twist" to his argument (msp. 7). However, there he is, computing a posterior without putting a prior on the parameters of the normal distribution. Apparently, he converts sampling distributions for estimators into posterior distributions for parameters. Isn't that fiducial inference?

According to King, our review of the "extended model" demonstrates error in Freedman et al. (1991). He does not explain the logic. Obviously, different neighborhoods in Los Angeles show different social characteristics—for both Hispanic and non-Hispanic inhabitants. That was true in 1991, and it is true today. What our review adds is this. If you know the answer, one of King's extended models may find it. But if you don't know the answer, the models are just shots in the dark.

Making the data available

King takes us to task for not providing data underlying our review. Although his other claims are mistaken, we did decline his request for data. His reaction seems disingenuous. After all, we had previously asked him for his data: he refused, sending us to the web. To read the files he pointed to, you need an HP workstation running UNIX and GAUSS. Even then, all you get is a long string of unidentified numbers. Apparently, the claim for replication on p. xix of King (1997) comes down to this: if you run his software on his files, with a platform of his choice, you will get his output.

It would be useful to have all the underlying data available in standard format (flat ASCII files, intelligibly documented). If King agrees to our plan and posts his data that way, we will post ours, along with the little simulation program used in Figure 1(c), and the version of EZIDOS that we used. That way, replication and independent analysis will be possible.

Summary and conclusions

King (1997) has a handful of data sets where his method succeeds. We have another handful where the method fails. Still other examples are contributed by Stoto (1998) and Tam (1998), with mixed results. Thus, King's method works in some data sets but not others. The diagnostics do not discriminate between probable successes and probable failures. That is the extent of the published empirical information regarding the validity of King's method. As a theoretical matter, inferring the behavior of subgroups from aggregate data is generally impossible: the relevant parameters are not identifiable. On this, there seems to be some agreement (Freedman et al., 1998, p. 1522; King, 1999). Thus, caution would seem to be in order—a characteristic not prominent in King (1997) or King (1999).

References

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, 1518–22.

Freedman, D. A., Klein, S. P., Sacks, J., Smyth, C. A., and Everett, C. G. (1991), "Rejoinder," *Evaluation Review,* 15, 800–816.

Goodman, L. (1953), "Ecological regression and the behavior of individuals," *American Sociological Review,* 18, 663–4.

Grofman, B. (1991), "Statistics without substance," *Evaluation Review,* 15, 746–69

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

King, G. (1998), A reply to Freedman et al.

Lichtman, A. (1991), "Passing the test," *Evaluation Review,* 15, 770-799

McCue, K. F. (1998), "Deconstructing King: Statistical Problems in *A Solution to the Ecological Inference Problem,*" Technical Report, California Institute of Technology, Pasadena.

Stoto, M. A. (1998), "A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data," Public Health Reports, 113, 182-83.

Tam, W. K. (1998), "Iff the assumption fits," *Political Analysis* 7, in press.