A REJOINDER TO BERK, BLALOCK, AND MASON

David A. Freedman*

We are all in Peter Marsden’s debt for organizing this discussion of important (and thorny) topics. Berk, Blalock, and Mason also deserve thanks for thoughtful contributions from different perspectives. There is substantial agreement among us on the state of empirical research in the social sciences. The disagreements—at least on diagnosis—seem relatively minor.

1. BERK

Berk and I see things the same way. Other than approval, I cannot add much to his comments. As is customary, my rejoinder will focus on points at issue.

2. BLALOCK

Blalock was the most critical. His position, if I understand him, can be summarized as follows:

1. The problems I discuss are very well known.
2. These problems are not confined to regression analysis.
3. Most empirical workers ignore such problems and are not doing a serious job of testing theory against reality.

This research was partially supported by NSF grant DMS 86-01634 and by the Miller Institute for Basic Research.

*University of California, Berkeley

353
4. Potentially useful statistical procedures quickly get converted to recipes and are then used quite thoughtlessly.

5. It is misleading for me to suggest "that experimental or quasi-experimental designs are presumably almost automatically superior to nonexperimental designs."

6. I should be kinder to sociologists because they are plagued by individual differences.

7. It is easier to pick the winners after the race has been run.

8. "We must make use of untested assumptions in all research, experimental or not."

I will respond to each point in turn.

1. Blalock may well have priority, but the state of the social science literature suggests that critical discussion of modeling issues is still in order.

2. Points 2, 3, and 4 are common ground.

5. Very little in our line of work is automatic. I don't agree with the details of his critique, but the bottom line is right: There are a lot of bad studies based on experiments and quasi experiments, just as Blalock says. On the other hand, there are a lot of good examples. My essay summarized one, and cited several others. With causal modeling in the social sciences, the situation is altogether different. Where are the real gains from the activity, practical or intellectual? There have been many public exchanges on the topic, and the one thing modelers do not provide is a bibliography of solid accomplishments. This omission is a strong signal.

6. I should be kinder, for other reasons. Individual differences are endemic, even in physics. A single molecule in a gas is quite unpredictable, and thermodynamic laws are statistical in nature. Epidemiology is more relevant; two examples illustrate the pervasive role of individual differences. First, as Snow's data show, most people who were exposed to contaminated water survived; some who drank clean water died. Still, cholera is an infectious, water-borne disease. Second, most smokers do not get lung cancer; some nonsmokers do. This was once an argument to show that cigarettes did not cause cancer. Even today, we cannot ex-
plain individual differences in susceptibility. Still, the evidence against cigarettes is rather strong.

7. Hindsight is 20-20, but this principle seems ungenerous when applied to Snow. An empirical test is almost possible: Spend an hour reading Snow’s book. You will see that he was a great scientific detective, and you will be richer for the experience.

8. Blalock and I seem to part company on the value of untestable assumptions. In my opinion, and others’, you’re not doing science unless you can test your theory. Path models were developed in part to make social science theories more testable, which lends considerable irony to the present discussion. However, I do not believe that modeling has advanced us very far towards the goal: t tests on regression coefficients don’t count for much. There seems to be considerable agreement on this point, and on the reasons.

In the end, of course, Blalock’s position on untestable assumptions comes close to mine. What is untestable with one data structure can be tested with another. He recommends (as I would)

that we must collectively make such assumptions explicit so that they can be readily challenged, and then rely on a cumulative process through which specific questionable assumptions are challenged, new data are collected to assess such assumptions, and models altered accordingly. (p. 332)

He goes on to say:

I therefore see no reasonable alternative to the development of much more complex causal models than we presently are accustomed to “testing.” I will fully admit, however, that we are presently failing rather miserably in engaging in such a collective effort. (pp. 332–33)

As to current use of regression models, Blalock is less forgiving than I am, and the two sentences just quoted are at the core of the issue. The similarity of his position to mine outweighs the differ-
ences. However, the optimism about complex models seems misplaced. We don’t test the current ones, and the next crop will be even more loosely connected to reality. The social science literature, viewed as a historical record, makes this point rather clearly.

Blalock’s explanation for our collective failure, in terms of the nine roadblocks to testing, is convincing. Our professional ethic rewards methodological innovation and technical elaboration. Institutionally, we are less appreciative of scholarship or the development of substantive knowledge. The consequences for our enterprise are bitter.

3. MASON

Mason’s points sound right to me, with one caveat. He writes:

At least since Durkheim and Weber, social scientists have debated whether quantification can be used to assess theories, models, ideas, views. In these debates, the difference between the evidence provided by experiments and nonexperiments . . . has played a minor role. Much more important have been issues of measurement, conceptualization, comparability, and validity. (p. 341)

Of course, the issues he cites are central. But so, I think, is the difference between experiments and nonexperiments. In crude but useful shorthand, regression models are popular because they seem to convert nonexperiments to experiments. That is how models derive cause from correlation. However, the derivation is legitimate only if the model’s assumptions about cause and effect are right. For additional discussion and references, see Lieberson (1985) or Freedman (1987) or the papers cited by Blalock.

The difficulty is a famous one. And it is handled as famous difficulties often are: acknowledged in a footnote and then ignored. But I think the issue is better faced now than later, and Mason agrees.

Mason has a series of hard questions for statisticians. I can respond only to the first one: Is the Bayesian approach likely to save the day for causal modeling? I think not. For thirty years, I have
found Bayesian statistics to be a rich source of mathematical questions. However, I no longer see it as the preferred way to do applied statistics, because I find that uncertainty can rarely be quantified as probability. The Reverend Thomas Bayes had his doubts too, which is why he allowed his essay to be published only after his death; and the matter has been debated ever since.

Mason goes on to say, “Statistics has definitely evolved into a field in which people can do their work without actually seeing and doing applications.” If anything, he is being tactful. Some members of the profession are trying hard to make changes, by teaching courses in which substantive questions come first and technique is introduced to find answers. Of course, all too often, technique comes first; data come in as purely decorative illustrations—a practice that is not confined to statistics departments.

4. CONCLUSION

Science grows out of the confrontation between theory and reality. Regression models seemed to offer good testing grounds for social science theory; but that promise has proved largely illusory. There is a surprising degree of agreement on this central point and on the reasons for the failure. It is much less clear, and naturally so, where we go from there.

Berk suggests a reduction in the level of aspirations, better designs, and cross-validation (including testing models on new data). This seems like good advice.

Blalock advocates more complex models and more rigorous testing. I wish him well, but I am not optimistic. If I am right, playing the game harder will not help. It is the rules that we need to change.

Mason recommends more use of experiments; although, as he notes, he will be using nonexperimental data for a long time to come. He writes, “Much, perhaps most, use of statistical inference in the social sciences is ritualistic and even irrelevant” (p. 343), and he recommends a change in journal editorial policies. Finally, he argues for better methodological training, with some help (to which he is surely entitled) from the statistics profession.

Where do I think we should go from here? My answer is fragmentary. Our situation is difficult but far from hopeless. We can retrieve it only by a long succession of small steps: Few can be taken
on auto-pilot, and some may turn out in the end to have been misdi-
rected. We should adopt the habit of making empirical claims that
are more sharply focused and perhaps more modest. We need to take
more seriously the job of comparing theory to reality. And we need
to build the requisite tools: reality tests instead of $t$ tests. It is not
complexity that will help us, but simplicity.

At any given time, most interesting questions will not have
empirical answers. Some do, however, and we have to identify them.
Then, different questions demand different kinds of answers. For
some issues, anecdotal evidence is the best that can be brought to
bear. For others, case studies are appropriate. At times, descriptive
statistics will help: $2 \times 2$ tables, or even a regression equation. A
formal statistical model with significance tests may be just the right
approach, on occasion. At present, such distinctions are seldom
made. And typical empirical papers, even the good ones, drift off
into fantasy. Yet the real world, with all its frustrations, is where we
belong.

REFERENCES

Lieberson, S. 1985. Making It Count: The Improvement of Social Theory and