



Better Rules for Better Decisions

Robert M. Hauser

Sociological Methodology, Vol. 25 (1995), 175-183.

Stable URL:

<http://links.jstor.org/sici?sici=0081-1750%281995%2925%3C175%3ABRFBD%3E2.0.CO%3B2-R>

Sociological Methodology is currently published by American Sociological Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/asa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

BETTER RULES FOR BETTER DECISIONS

*Robert M. Hauser**

About a decade ago, after David Grusky and I had suffered endlessly over the choices among alternative models in our comparative analyses of social mobility (Grusky and Hauser 1984), our satisfaction with the product of those analyses was temporarily shattered by the news that Adrian Raftery (1986) would publish a methodological comment on the work. What would he have to say? And could we defend our work? Would we take the standard defensive posture of sociologists whose work was under criticism?

In actuality, Raftery's brief and elegant comment turned out not to require a defense at all. Rather, it outlined a superior way to think about the decisions that we had faced—namely, how to choose among alternative models in a sample so large that standard inferential methods would lead us to reject all but a saturated model. Raftery's proposal to use the Bayesian information criterion (BIC) relieved, rather than increased our discomfort at having ignored standard rules of statistical inference, and it even supported some—though not all—of the decisions that we had made.¹

Pleased as I was that some parts of the Grusky-Hauser analysis survived Raftery's scrutiny, I am happier yet that our efforts prompted the introduction of a simple and defensible rule of thumb that could be used to improve decisions in discrete multivariate analysis and structural equation models (Raftery 1993). For the past several years, I have routinely used BIC as a guide in model selection (Hauser and Wong 1989; Wong and Hauser 1992; Hout and Hauser 1992; Hauser 1993; Hauser and Phang 1993; Kuo and Hauser 1995a, 1995b), sometimes without showing the details of inferential procedures in the text.

*University of Wisconsin, Madison

¹At the time, as I recall, we even wanted to write a "reply" to the comment that would express our agreement and gratitude to Raftery, but my recollection is that the editor discouraged this unexciting, though original, use of journal pages.

Raftery has now explained and elaborated the uses of Bayesian inference in a wider context. In this brief commentary, I offer a few additional examples and suggestions about the use of Bayesian methods of model selection, as well as encourage readers to read, evaluate, and use Raftery's contributions in their own teaching and research.

1. FURTHER REFLECTIONS ON COMPARATIVE SOCIAL MOBILITY

I will start by adding a footnote to Raftery's exposition of the right way to do the Grusky-Hauser analysis. In the early 1980s, the central idea in cross-national mobility analysis was the so-called Featherman-Jones-Hauser (FJH) hypothesis (Featherman, Jones, and Hauser 1975). It said that once marginal distributions were controlled (in a log-linear model), the interactions (odds-ratios) in a mobility table were essentially invariant across societies. Grusky and Hauser wanted to show that mobility regimes in different countries were similar, but not identical, in order to motivate their effort to explain the cross-national variations in interactions without entirely denying the usefulness of the FJH hypothesis. Thus, among other models, they estimated a variant of the quasi-symmetry (quasi-perfect mobility) model, shown in line 3 of Raftery's Table 2, and a variant of the saturated model, shown in line 4 of Raftery's Table 2. Both models specified that the interaction effects were the same in all 17 countries. Each model yielded a nominally significant likelihood-ratio test statistic, relative to its less constrained counterpart (1350 with 45 df and 1329 with 60 df, respectively).

In the article, Grusky and Hauser (1984, p. 26) waffled about this, first noting that these contrasts were "highly significant statistically," then saying that "The results . . . imply convergence . . ." (both among industrialized countries and in the full sample of 17 countries), and, finally, going on to model cross-national differences. Their analysis eventually yielded the explanatory model whose satisfactory fit is reported in line 5 of Raftery's Table 2. Had Grusky and Hauser used BIC, they would have been able to accept the model of cross-nationally variable quasi-symmetry, but they would also have had to reject outright both the model of constant quasi-symmetry (BIC = 826) and that of constant association (BIC = 631). Grusky and Hauser did go on in the right direction, but not for quite the right reason. Had they been more forthright in this

early part of the history of the FJH hypothesis, it is possible that others would not later have gone to such great lengths to sustain it (Erikson and Goldthorpe 1992).

What is the moral of this story? The first thing most researchers will notice about the decision rules that Raftery elaborates is that they are far more conservative, at least in survey-based samples, than application of the customary criteria of $P = 0.05$ or $P = 0.01$. In my experience, people are uncomfortable with this kind of conservatism—with or without the legitimation of the BIC approximation—even though its rewards are likely to include both scientific parsimony and external validity. I like this elaboration of Raftery's example because it shows how a conservative but informal decision rule can also go wrong. In this case, BIC not only tells us which "significant" findings should be ignored but which should have been pursued most seriously.

2. ASSESSING THE REPLICABILITY OF FINDINGS ON OCCUPATIONAL SCALING

Raftery's text several times notes the value of BIC in improving "out-of-sample predictions," that is, in yielding replicable findings. This phrase suggested another application of BIC—namely, an analysis in which the investigators could not determine the expected value of a new measure of correlation.

In the May 1992 *American Journal of Sociology* there was an extensive symposium on the scaling of occupational status. Rytina (1992a) introduced a method for constrained (symmetric) canonical scoring of the rows and columns of an intergenerational social mobility table, the Symmetric Scaling of Intergenerational Continuity (SSIC). He estimated scores for a 308 by 308 occupational mobility table (from the General Social Survey [GSS], 1972 to 1986, $n = 7965$) and obtained a much higher correlation between father's and adult child's occupation (0.450) than was obtained with the Stevens-Featherman variant of the Duncan SEI (0.338). He concluded that the intergenerational correlation was truly higher and that the SEI failed to reflect a large share of the vertical dimension of occupational stratification.

Hauser and Logan (1992) responded by cross-validation and by example. First, in a fresh sample (the GSS from 1987 to 1990) Rytina's SSIC scores (from the 1972 to 1986 data) were less highly

correlated across generations than was the SEI. Second, arguing by analogy, they showed that the canonical correlation (not that based on the SSIC), when corrected for loss of degrees of freedom using a formula of Lawley (1959), was of about the same magnitude as the observed correlation based on the SEI.² They were unable to come up with an analytic correction of the correlation based on the SSIC. Rytina (1992b) rejoined at length, refusing to accept the idea that sampling variability increased the SSIC correlation.

Raftery's paper suggests a simple analysis that should resolve the issue: Compute BIC' for the two competing models (from his equation 26). With $N = 7965$, $p_1 = 1$, and $r_1 = 0.338$, $BIC' = -957$. With the same sample size, but $p_2 = 307$ and $r_2 = 0.450$, $BIC' = 955$.³ In other words, Rytina's model fares worse than no model at all. It would be superior to that based on the SEI if it could produce a correlation as large as 0.450 with 92 or fewer parameters. That would yield $BIC' = -976$. Any larger number of parameters would yield a difference in BIC' (relative to the SEI equation) less than 10 (in absolute value). Obversely, using as many as 307 parameters, we would have to obtain a correlation of 0.62 or more to improve on what the SEI does with one parameter. In short, this simple analysis would lead us to expect the failure of cross-validation that was observed by Hauser and Logan.⁴

3. SELECTION OF VARIABLES IN SINGLE-EQUATION MODELS

When I first read an earlier draft of Raftery's paper, I was put off slightly by his emphasis on the problem of variable selection in

²The idea is the same as that of correcting a multiple correlation for its degrees of freedom.

³This illustrates a use of Raftery's equation 26 that he did not elaborate—namely, to evaluate differences in overall fit (R^2) between linear models that differ in more than one degree of freedom. The application to canonical scoring is a bit unusual; a more common application will be to tests of the effects of sets of categories, as in dummy-variable regression analysis.

⁴Hauser and Logan (1992, p. 1694) introduced a design effect of two-thirds for the GSS sample, and its use here would yield an even stronger rejection of Rytina's finding. In a sample two-thirds as large as the cumulative GSS ($n = 5310$), using 307 parameters, we would have to obtain an SSIC correlation of 0.68 to compete with that based on the SEI, or, obversely, we would have to obtain an SSIC correlation of 0.45 with only 64 parameters.

single-equation models. As described above, my introduction to BIC took place in the context of alternative specifications of models for cross-classification tables. My research of late has focused on constraints on slopes, means, and variances within and between populations in structural equation models. Why should we worry so much about inference in single-equation models?

First, the message comes through loud and clear that Raftery's main ideas apply as well to these other modeling situations as to single-equation models, even if the machinery for identifying the sets in Occam's window is not yet available. One most valuable suggestion of Raftery's paper in this respect is that standard software for structural equation models should extend displays of *t*-statistics and modification indices to approximate BIC (as well as reporting BIC along with other standard indices of model fit). Thus, as he notes, from estimation of a single model, we can easily approximate Bayes factors both for the deletion of every parameter in the model and for the release of every constrained parameter of the model.

Second, despite my interest in other aspects of structure, the more I think about it, the more I think sociologists can profit from better methods for the selection of variables in relatively simple models. In this context, Raftery's illustrative analysis of the problem of model uncertainty in the data on crime and punishment deserves close and repeated reading. It provides a powerful example of how we ought to proceed when our interest lies in the full set of variables that belong (or do not belong) in an equation.

For example, Jencks, Perman, and Rainwater (1988) introduced a new index of job desirability (IJD), which, they argued, might be preferable to occupational status or earnings in analyses of social stratification. To construct the index, they used a magnitude scaling item to measure the desirability of the job held by each adult in a national Survey of Job Characteristics (SJC, $N = 809$). Then they regressed the desirability of the job on a vector of 48 job characteristics, eliminating all but the 14 that proved statistically significant.⁵ Their proposal is that rather than ascertaining occupation and mapping it into a status or prestige scale, future researchers should

⁵Jencks, Perman, and Rainwater (1988 pp. 1336–37) were not unmindful of the problem of replication. In fact, they estimated the reliability of the IJD using independent half-samples. However, they do not appear to have cross-validated their selection of regressors.

simply ascertain the significant job characteristics and weight them as they affected job desirability in the SJC.

Which and how many job characteristics would have appeared in the IJD if Jencks, Perman, and Rainwater had used Bayesian methods of model selection? A full analysis would require location of Occam's window, but it is instructive to note that to meet Raftery's criterion of "very strong" evidence for inclusion ($BIC = 10$), a variable should have a t -statistic greater than 4.09 in a sample of 809. This criterion is met by only two of the variables in the published, 14-variable version of the IJD equation, the earnings and the educational requirements of the job (p. 1336).⁶ Only five of the 14 variables meet Raftery's less stringent criterion of "strong" evidence. Perhaps further analytic work could identify a small enough set of valid predictors of the IJD to warrant measurement of those components on a regular basis, even outside specialized studies of social stratification.

4. INTRODUCING BIC IN RESEARCH AND TEACHING

It is fortunate that we now have a lucid, complete, and relatively nontechnical guide to the use of BIC and BIC'. It is not easy to break old habits or to teach students to develop the right habits in the conduct of research. In the fall of 1994 I introduced BIC in my course in structural equation models, and it was often difficult and in some cases impossible to persuade students not to reject a null hypothesis, especially one pertaining to the overall fit of a model, merely because the deviance was nominally statistically significant with $P = 0.05$ or $P = 0.01$. We want so badly to reject null hypotheses that we often do so badly—and often, thanks to the miracles of modern software—without much regard for the theoretical ideas that we actually wish to test. Perhaps one inducement to good practice will be Raftery's finding that, just as we set nominal probability levels too high in large samples, we also set them too low in small samples.

While I hope that Raftery's paper will lead readers to appreciate the meaning of Bayes factors, in practice it may be difficult even to introduce the rules of thumb suggested by Raftery in his Tables 6, 7, 8, and 9. Perhaps it will help to photocopy those tables and post

⁶It is striking that these two job-level variables are analogs of the components of occupational socioeconomic status.

them visibly in statistical laboratories. They ought to be circulated by journal editors along with guidelines to authors. They ought to become as commonplace as tables of the distribution of t or of χ^2 . The formulas on which the tables are based are so simple that I would encourage readers to validate the tables in a computer spreadsheet. Having done that, the spreadsheet will also estimate critical values of BIC for the sample at hand, and it can be used to calculate BIC from standard computer output.⁷

Another good way to introduce Bayesian selection methods will be to encourage students (and researchers) to compare standard model selection procedures with the use of BIC in simulated, null data, or repeated samples. Those who lack theoretical skills in statistics may well be able to carry out their own analyses of simulated data, just as described by Raftery. Another straightforward exercise, accessible to any analyst, will be to combine half-sample replication with the use of BIC and other model selection procedures. For example, split a sample into two random halves. Run a regression equation in one half-sample, and choose which variables should enter the equation using several decision rules—for example, all significant variables in the full equation, forward selection, backward selection, and BIC'. Then estimate the same equation in the second half-sample and compare the pattern of significant coefficients with those determined in the first-half sample under varying decision rules. Such exercises ought to enter the statistical curriculum.

Over the past 20 years, the frontier of sociological modeling has advanced rapidly, but with few exceptions, there has been little progress in our practice of statistical inference. To Raftery's credit, he makes no grandiose claims or promises about the value of Bayesian methods, and he also emphasizes the importance of other theoretical and methodological aspects of the research process. This is good as well as modest advice, and I second Raftery's observation that there are no methodological panaceas. At the same time, I hope and expect that Raftery's exposition of Bayesian methods of model selection represents the beginning of their widespread use in sociological research as well as an invitation to further advances in inferential methods. Raftery's methodological work, like that of Leo Good-

⁷For many years, I have constructed statistical tables by importing standard output into a spreadsheet and using the spreadsheet both to format the table and to perform and document auxiliary calculations.

man, is truly useful to researchers, and in use it will repay close and repeated study.

REFERENCES

- Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- Featherman, David L., F. Lancaster Jones, and Robert M. Hauser. 1975. "Assumptions of Social Mobility Research in the U.S.: The Case of Occupational Status." *Social Science Research* 4 (December):329-60.
- Grusky, David B., and Robert M. Hauser. 1984. "Comparative Social Mobility Revisited: Models of Convergence and Divergence in 16 Countries." *American Sociological Review* 49 (February):19-38.
- Hauser, Robert M. 1993. "Trends in College Entry Among Blacks, Hispanics, and Whites." In *Studies of Supply and Demand in Higher Education*, edited by Charles Clotfelter and Michael Rothschild. National Bureau of Economic Research. Chicago: University of Chicago Press.
- Hauser, Robert M., and John A. Logan. 1992. "How Not to Measure Intergenerational Occupational Persistence." *American Journal of Sociology* 97(6)(May):1689-711.
- Hauser, Robert M., and Hanam Samuel Phang. 1993. "Trends in High School Dropout Among White, Black, and Hispanic Youth, 1973 to 1989." IRP discussion paper, vol. 1007-93. Madison, Wis.: Institute for Research on Poverty.
- Hauser, Robert M., and Raymond Sin-Kwok Wong. 1989. "Sibling Resemblance and Inter-sibling Effects in Educational Attainment." *Sociology of Education* 62 (July):149-71.
- Hout, Michael, and Robert M. Hauser. 1992. "Symmetry and Hierarchy in Occupational Mobility: A Methodological Analysis of the CASMIN Model of Class Mobility." *European Sociological Review* 8 (December):239-66.
- Jencks, Christopher S., Lauri Perman, and Lee Rainwater. 1988. "What Is a Good Job? A New Measure of Labor Market Success." *American Journal of Sociology* 93(6)(May):1322-57.
- Kuo, Hsiang-Hui Daphne, and Robert M. Hauser. 1995a. "Gender, Family Configuration, and the Effect of Family Background on Educational Attainment." *Social Biology*, forthcoming.
- . 1995b. "Black-White Differentials in the Effect of Family Background on Educational Attainment of U.S. Men Born from 1907 to 1946: A Study of Sibling Resemblance." *Sociology of Education* 68(2) (April):136-60.
- Lawley, D. N. 1959. "Tests of Significance in Canonical Analysis." *Biometrika* 46:59-66.
- Raftery, Adrian E. 1986. "Choosing Models for Cross-Classifications (Comment on Grusky and Hauser)." *American Sociological Review* 51(1)(February):145-46.
- . 1993. "Bayesian Model Selection in Structural Equation Models." In

- Testing Structural Equation Models*, edited by Kenneth A. Bollen and J. Scott Long, 163–80. Newbury Park Calif.: Sage Publications.
- Rytina, Steve. 1992a. "Scaling the Intergenerational Continuity of Occupation: Is Occupational Inheritance Ascriptive After All?" *American Journal of Sociology* 97(6)(May):1658–88.
- . 1992b. "Response to Hauser and Logan and Grusky and Van Rompaey." *American Journal of Sociology* 97(6)(May):1729–48.
- Wong, Raymond Sin-Kwok, and Robert M. Hauser. 1992. "Trends in Occupational Mobility in Hungary Under Socialism." *Social Science Research* 21 (4)(December): 419–44.