

- Young, Cristobal. 2009. "Model Uncertainty in Sociological Research: An Application to Religion and Economic Growth." *American Sociological Review* 74:380–97.
- Young, Cristobal. Forthcoming. "Model Uncertainty and the Crisis in Science." *Socius*.
- Young, Cristobal, and Katherine Holsteen. 2017. "Model Uncertainty and Robustness: A Computational Framework for Multi-model Analysis." *Sociological Methods and Research* 46:3–40.

Author Biography

Robert M. O'Brien is a professor emeritus at the University of Oregon. He specializes in criminology and quantitative methods. Within criminology, he focuses on the methods used to gather criminological data, on the analysis of crime rates, and on the task of extricating the effects of ages, periods, and cohorts on crime rates. His most recent publication on that topic, "Homicide Arrest Rate Trends in the United States: The Contributions of Periods and Cohorts (1965–2015)," appeared in 2018 in the *Journal of Quantitative Criminology*. In quantitative methods, some of his contributions involve the effects of using interval data as ordinal, generalizability theory, identification in structural equation modeling measurement models, the use of multicollinearity indices, and an obsession with age-period-cohort models. In 2015 he published a book on this topic, *Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data* (Chapman & Hall, 2015).

COMMENT: BAYES, MODEL UNCERTAINTY, AND LEARNING FROM DATA

*Bruce Western**

*Columbia University, New York, NY, USA

Corresponding Author: Bruce Western, bruce.western@columbia.edu

DOI: 10.1177/0081175018799095

I was looking at some regression results with a research assistant last week. The RA had been preparing administrative data from a jail, and I was interested in the correlates of returning to incarceration among those who were released. The big challenge in this analysis was simply cleaning up administrative records that were not generated for statistical analysis. Data cleaning had taken several months. We were eagerly awaiting these first regressions, and I wanted to look at basic demographic patterns: race and sex differences in reincarceration. The first regressions we ran also controlled for age, criminal offense, and county of jurisdiction. The initial output indicated that blacks were more likely to go back

to jail than whites and that men were more likely to return to custody than women. Unexpectedly, the results indicated that older people were more likely to return to jail than younger people. This contradicted the age-crime curve, the universal finding that criminal offending declines with age. Hmm.

I asked the RA to pull up the coding of the age variable, but in the meantime I was thinking about why people older than 40 were more likely to return to incarceration than those in their 20s. Jails, more than prisons, incarcerate people who face a lot of social and economic insecurity. Problems of homelessness and mental illness are highly prevalent. Perhaps enduring needs for drug treatment and housing among older people explained their high rates of reincarceration. Just then, my RA announced that she had found the problem with the age code and fixed it. We fit the regression again. Now, the rate of reincarceration declined smoothly with age, just as expected.

A lot of data analysis looks like this. We begin with some intuitions about what patterns we will see in the data. But often, expectations are fairly weak, and if the data point in a different direction, we sort through an infinite catalog of *post hoc* explanations to make sense of the results. A lot of sociological data analysis does not involve the testing of a well-specified null hypothesis, but it is an iterative and inductive process of learning from data. Prior information is diffuse, and we readily invent explanations for unusual findings, at least for a few minutes in the privacy of our offices.

The article by Muñoz and Young (this volume, pp. 1–33) offers one approach to the problem of learning from data in a context with weak prior information. They consider model uncertainty in the familiar applied situation in which interest focuses on one predictor in a regression analysis, but the researcher is unsure which control variables to include in the model.

More formally, for a regression on the vector y with a predictor of interest, x ,

$$y = \beta x + Z' \gamma + \varepsilon,$$

Z is a matrix of covariates that includes a column of ones for the intercept, and ε is a residual. The matrix, Z , is uncertain in the sense that the covariates that constitute the columns of Z are not known. Researchers manage model uncertainty by experimenting with different covariates. Point estimates of β change from model to model, depending on the correlation between x and Z , and nominal p values are incorrect.

The authors address the problem of model uncertainty with a diagnostic in which all possible K covariates are specified *a priori*, yielding 2^K possible models and a corresponding set of estimates, $\hat{\beta}_k$ ($k = 1, \dots, K$).

Given a preferred estimate, $\hat{\beta}_p$, the authors propose a measure of model sensitivity they call a “robustness ratio,” $\hat{\beta}_p/s$, where s is the standard deviation of the estimates, $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K$. (The statistic is defined in Young and Holsteen [2017:13].) When the robustness ratio is less than 2, they argue that inference about β should be treated as highly sensitive to the model specification. On the basis of simulation experiments with $\beta = 0$, they suggest that accepting the null hypothesis when the robustness ratio is less than 2 improves on conventional inference with nominal p values and reduces the rate of false positives, of falsely concluding that β is nonzero.

The problem of model uncertainty is a fundamental applied challenge in quantitative sociology. The authors’ language of false positives is reminiscent of Bonferroni adjustments and the frequentist analysis of multiple independent comparisons, but the distinct problem of model uncertainty has been fully formalized from a Bayesian perspective. In the notation above, we have K discrete candidate models, M_k , each yielding a posterior inference for the coefficient of interest, $p(\beta|y, M_k)$, where we write the posterior distribution as conditional both on the data, y , and the model, M_k . Unconditional inference for the regression coefficient is obtained by averaging over models,

$$p(\beta|y) = \sum_{k=1}^K \pi_k p(\beta|y, M_k),$$

where π_k is the posterior probability of model, M_k . Raftery (1996) showed that with diffuse prior information for the regression coefficients, the posterior model probability can be approximated by a function of the Bayesian information criterion. The posterior expectation of β , averaging over model uncertainty, is

$$E(\beta|y) \equiv \hat{\beta} = \sum_{k=1}^K \pi_k E(\beta|y, M_k),$$

where $E(\beta|y, M_k)$ is the posterior mean of β under model k and might be estimated by $\hat{\beta}_k$. The posterior variance averaging over model uncertainty is given by

$$V(\beta|y) = \sum_{k=1}^K \pi_k V_k + \sum_{k=1}^K \pi_k (\hat{\beta}_k - \hat{\beta})^2,$$

where V_k is the posterior variance under model k . The Bayesian analysis formalizes Muñoz and Young’s idea of a “total standard error” in which the unconditional posterior variance has two components: a within-model variance, $\sum_k \pi_k V_k$, and a between-model variance, $\sum_k \pi_k (\hat{\beta}_k - \hat{\beta})^2$. This formalization and computational formula are detailed by Draper

(1995) and Raftery (1996). The all-possible subsets approach described by the authors was pioneered in the classic article by Leamer (1983).

The Bayesian framework provides a few extensions to the analysis by Muñoz and Young. First, even when a model is robust according to the authors' diagnostic, standard inference for a preferred model will still yield standard errors that are too small. This is because in the authors' approach, once a model is judged to be robust, inference is conditional on a single preferred specification instead of averaging over all possible models. Second, the authors' idea of a total standard error effectively places equal probability on every point in the model space, and the data themselves are not informative about the probability of each model; for Bayes, the data are informative about the probability of different models and greater weight is given to more probable models. Third, the instability of coefficients across models varies not just with sample size but also collinearity. The authors' simulations illustrate some intuitions about model sensitivity and sample size, but only in the unrealistic situation where covariates are, on average, uncorrelated.

Nevertheless, drawing attention to the problem of model uncertainty, the pitfalls of uncalibrated data mining, and the inflation of nominal p values is unquestionably important. The key texts cited by Muñoz and Young are Leamer (1983) and Freedman (1983) and should be required reading in every graduate methods sequence in sociology.

Some readers may feel that the authors' motivating example of inference about a treatment effect with observational data is anachronistic. Certainly, the limits of observational data for learning about causal effects are more widely understood now than in 1983, when Leamer was first taking the con out of econometrics. Indeed, the authors' analysis of the job-training experiment illustrates how a strong experimental design can greatly reduce model uncertainty. In any case, the problem of statistical inference for which the issue of model uncertainty is relevant is distinct from the problem of inference about a causal effect.

The analysis of model uncertainty is one part of a larger methodology of data analysis that provides rules for learning from data, beyond estimation and inference for particular models. In the methodology of data analysis, I would also include the statistics of data visualization and exploration, outlier detection and other regression diagnostics, and more informal inquiries such as placebo tests and alternative codings of variables. For some methods, the objective is not statistical inference but description. Given a data description, other methods are concerned with sensitivity: assessing the dependence of a description on the specific set of observations and the underlying assumptions.

The problem of statistical inference can occupy an outsized place in quantitative sociology, where researchers work under conditions of

radical model uncertainty—where even the age-crime curve can be doubted. In these settings, our largest contributions may involve making new observations with help from a rigorous research design, or developing new descriptions of secondary data.

Acknowledgment

I would like to thank Nate Wilmers, who provided comments on an earlier draft.

References

- Draper, David. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society, Series B* 57:45–97.
- Freedman, David A. 1983. "A Note on Screening Regression Equations." *American Statistician* 37:152–55.
- Leamer, Edward E. 1983. "Let's Take the Con out of Econometrics." *American Economic Review* 73:31–43.
- Raftery, Adrian E. 1996. "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalised Linear Models." *Biometrika* 83:251–66.
- Young, Cristobal, and Katherine Holsteen. 2017. "Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis." *Sociological Methods and Research* 46:3–40.

Author Biography

Bruce Western is a professor of sociology and codirector of the Justice Lab at Columbia University. His current research interests focus on methods of data collection and analysis for hard-to-reach populations and the institutional contexts of American racial inequality and poverty.

REJOINDER: CAN WE WEIGHT MODELS BY THEIR PROBABILITY OF BEING TRUE?

*John Muñoz**
Cristobal Young†

*Stanford University, Stanford, CA, USA

†Cornell University, Ithaca, NY, USA

Corresponding Author: John Muñoz, jdmunoz@stanford.edu

DOI: 10.1177/0081175018796841