

A Unified User-Interface for Single and Multi-Equation Models (Also Known As, “Zelig: Everyone’s Statistical Software”)¹

*Kosuke Imai*² *Gary King*³ *Olivia Lau*⁴

Abstract

We propose a common framework for statistical analysis and software development built on and within the R language. Researchers in different academic disciplines have developed different statistical models, different mathematical notation, different parameterizations, different quantities of interest, and hence different computational implementations. The users of statistical software have much to gain by navigating the babel of R’s many packages, but it is often far more difficult than it should be, given that packages have so much underlying statistical theory and computational structure in common. To address this problem, we have developed a conceptual framework and software package that offers:

- A common syntax for specifying univariate response, multivariate response, multilevel, and hierarchical models. In particular, we offer a user-specified, intuitive interface for identifying constraints across equations.
- A method to use this syntax with existing R packages, without modifying those packages (by re-defining function calls on the fly).
- A method to translate this syntax into matrices and arrays useful for programmers. For multiple equation models, we offer three implementation options (all of which work with user-specified constraints): an intuitive option that stacks matrices visually, a computationally efficient option that creates arrays of explanatory variables, and a memory-efficient option that coerces parameters to matrices.
- A framework for calculating quantities of interest with or without conditioning on the observed values of a particular unit, to provide users with substantive interpretation of model output.
- Methods to process lists of multiply imputed data, or stratified data, and combine model estimates when generating quantities of interest.
- An application programmer interface that makes it possible to dynamically generate a graphical user interface (GUI) for the models included in Zelig (see, e.g., the Virtual Data Center, for one example of a GUI which has already been implemented).

Our approach is intended to bring the computational flexibility and power of R to users who need to understand the data, the modeling assumptions, and the quantities of interest, but are not as focused on computational algorithms as applied statisticians. By creating a common interface between users and developers, and a set of tools to operate on that interface, we hope to extend the range of models that one can write and use in R, and the range of people to whom R is accessible.

¹Our thanks to the National Institutes of Aging (P01 AG17625-01), the National Science Foundation (SES-0318275, IIS-9874747), and the Mexican Ministry of Health for research support. Current software is available from CRAN or <http://gking.harvard.edu/zelig/>.

²Assistant Professor, Department of Politics, Princeton University, kimai@princeton.edu

³David Florence Professor of Government, Department of Government, Harvard University, king@harvard.edu

⁴Ph.D. Candidate, Department of Government, and M.A. Student, Department of Statistics, Harvard University, olau@fas.harvard.edu