A Backbone for Simulating Quantities of Interest from Generalized Linear Models

by Christopher Gandrud

Abstract Simulating quantities of interest estimated from generalized linear models (GLM) can be an effective way to explore and communicate statistical results. **coreSim** provides core functions that can serve as a reliable and extensible "backbone" to new packages for simulating quantaties of interest for finding and plotting simulated quantities of interest from GLMs. I demonstrate how to use **coreSim** as a backbone by showing how it is incorporated into the new **pltesim** package for simulating and displaying probabilistic long-term effects in models with temporal dependence.

Introduction

There has been a recent trend in the social sciences to improve the interpretation of results from generalized linear models (GLM) by presenting estimated quantities of interest from substantively meaningful scenarios (e.g. Gandrud, 2015; Licht, 2011; Williams and Whitten, 2012). King et al. (2000) advanced a post-estimation simulation technique for estimating the uncertainty around these estimates. A number of R packages implement this approach [CITE]. In particular, the **Zelig** package (Choirat et al., 2016) implements this technique for a wide range of model types in an attempt to be "everyone's statistical software".

While all of these packages based on the same simulation procedure, all of these packages rely on different, custom built ways of doing these simulations. This presents unnecessarily increases the labor needed to use post-estimation simulations for showing results in new modelling situations. Current implementations of these packages have also proven to be unreliable in ways directly related to their architecture.

In this paper I introduce a new R package–**coreSim**–that aims to be an extensible and reliable "backbone" for future analyses and R packages that find post-estimation simulations for substantively relevant scenarios. I give an example of this by showing how **coreSim** forms the basis of the new **pltesim** package for simulating and displaying probabilistic long-term effects in models with temporal dependence.

Post-estimation simulations from GLMs

One way to estimate substantively meaningful quantities of interest for complex scenarios from with their associated uncertainty is through post-estimation simulations.[1] The procedure is straightforward. Remember that a generalized linear models can be summarized by two equations. One describes the stochastic component:

$$Y_i \sim f(\theta_i, \alpha) \tag{1}$$

Here the dependent variable Y is a random drawn from the probability density function $f(\theta_i, \alpha)$. The other equation describes the systematic component:

$$\theta_i = g(X_i, \beta), \tag{2}$$

where X is a vector of explanatory variables and β is a vector of effect parameters. The function g is often referred to as the link function, which specifies how the variables and effect parameters are translated into θ.

We can use post-estimation simulations to find and communicate substantively meaningful results from these models. To do this we first estimate $\hat{\beta}$ and $\hat{\alpha}$. Second, we drawn n number of values of these parameters from the multivariate normal distribution using the parameter point estimates and their variance covariance matrix with a mean of $\hat{\gamma}$–a vector created by stacking $\hat{\beta}$ and $\hat{\alpha}$–and variance of $V(\hat{\gamma})$:

This is straightforward using two functions from the **stats** package–`coef` and `vcov`–as well as the `mvrnorm` function from the **MASS** package. Both **stats** and **MASS** are included with the default R installation. For example:

[1] See King et al. (2000) for a comparision with related fully Bayesisan Markov chain Monte Carlo and Bootstrapping methods.

```
# Load package that contains the Prestige data set
library(car)

# Estimate normal linear model
m1 <- lm(prestige ~ education + type, data = Prestige)

# Extract
```

Previous post-estimation simulation packages

The **Zelig** package could potentially help streamline this process and in so doing form the basis of R packages that generate post-estimation simuliations in novel areas.

Zelig has a number of issues that have made it unreliable over time. By aiming to be everyone's statistical software, **Zelig** relies on many (13 imported) dependencies.

Summary

Bibliography

C. Choirat, J. Honaker, K. Imai, G. King, and O. Lau. *Zelig: Everyone's Statistical Software*, 2016. URL http://zeligproject.org/. Version 5.0-12. [p1]

C. Gandrud. simPH: An R package for illustrating estimates from cox proportional hazard models including for interactive and nonlinear effects. *Journal of Statistical Software*, 65(3):1–20, 2015. URL http://www.jstatsoft.org/v65/i03/. [p1]

G. King, M. Tomz, and J. Wittenberg. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*, 44(2):347–361, 2000. [p1]

A. A. Licht. Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis. *Political Analysis*, 19:227–243, Sept. 2011. [p1]

L. K. Williams and G. D. Whitten. But Wait, There's More! Maximizing Substantive Inferences from TSCS Models. *Journal of Politics*, 74(03):685–693, 2012. [p1]

Christopher Gandrud
City University London and Hertie School of Governance
Rhind Building
London, EC1V 0HB
christopher.gandrud@city.ac.uk