Online Appendix to "A Bootstrap Method for Conducting Statistical Inference with Clustered Data"

Jeffrey J. Harden*

Abstract

This online appendix contains four parts: a discussion on implementing BCSE in applied work, code for estimating BCSE in `R` and Stata, a comparison between BCSE and the CLRT method shown in Erikson, Pinto, and Rader (2010), and a replication of the Brown, Jackson, and Wright (1999) model with CLRT.

1 Implementing BCSE in Applied Work

The main drawback to using the BCSE method is that calculating the standard errors for a model multiple times will produce different estimates each time because different bootstrap samples are drawn each time. This leads to the question of how to report results in journal articles and presentations. This is an issue with any simulation-based technique, such as conventional bootstrapping or Markov Chain Monte Carlo (MCMC) methods. The important points to note are that the analyst controls the number of simulations and adding more simulations brings the estimate closer to the true value. In the case of BCSE, increasing the number of bootstrap replications (B) shrinks the variation between calculations.[1] From this, the analyst could choose the number of digits to round off from each standard error estimate when reporting results, and increase B until multiple calculations return the same rounded value.[2] In addition, the seed of the random number generator can be set to duplicate the same bootstrap samples in later iterations.

2 BCSE in Software

The following code provides examples for estimating BCSE and calculating ρ in `R` and Stata. In each one, the dependent variable `y` is regressed on independent variables `x1`, `x2`, and `x3`, ob-

*Ph.D Candidate, Department of Political Science, University of North Carolina at Chapel Hill, 312 Hamilton Hall, CB #3265, Chapel Hill, NC 27599, jjharden@unc.edu

[1]Feng, McLerran, and Grizzle (1996) recommend using at least 500 bootstrap replications.

[2]For example, if the analyst wishes to report three digits of each standard error, B could be increased until multiple calculations return the same value at the third decimal place.

servations are clustered in the variable `cluster`, and B is set to 1,000. This code is for an OLS model, but the process is similar for other models, such as logit, probit, or duration analyses.

2.1 Example R Code

```
library(Design)
library(Hmisc)
model <- ols(y ~ x1 + x2 + x3, x = TRUE, y = TRUE)
BCSE <- bootcov(fit = model, cluster = cluster, B = 1000)
rho <- deff(y = model$residuals, cluster = cluster)
print(BCSE); print(rho)
```

2.2 Example Stata Code

```
* Method 1
regress y x1 x2 x3, vce(bootstrap, rep(1000) cluster(cluster))
* Method 2
bootstrap, rep(1000) cluster(cluster): regress y x1 x2 x3
predict residuals, residuals
loneway residuals cluster
```

3 BCSE and CLRT

In the article, I identify some differences that make BCSE preferable to the CLRT method: (1) CLRT does not calculate a full covariance matrix of the parameter estimates, (2) CLRT is driven by a much different philosophy than the other methods I examine, and (3) CLRT is more difficult to implement than BCSE. Below I discuss each of these issues.

3.1 Covariance Matrix vs. Hypothesis Testing

Randomization testing was designed specifically to get a p-value for a statistic of interest (Fisher 1922). This is different from estimation of a covariance matrix (and thus, standard errors), which is a complete representation of model uncertainty that can be used to get a p-value. Erikson, Pinto, and Rader (2010) obtain standard errors out of their CLRT procedure by calculating the variance of the resampled estimates, but the method does not calculate covariance between coefficients. This is problematic because a covariance estimate is often needed for analysis of model results. For instance, conducting joint F-tests, calculating confidence intervals for the marginal effects of variables in interaction models (e.g., Figure 6 of my article) and the commonly used CLARIFY procedure of King, Tomz, and Wittenberg (2000) all require an estimate of covariance between coefficients of interest. The BCSE method provides an estimate of the full covariance matrix.

3.2 Philosophy

CLRT also requires researchers to make a substantial shift in how they view error in their models. The general method of randomization testing conceives of the stochastic element of a model as error fixed (in repeated samples) to observations. This differs from the traditional econometric view that the stochastic element is additive error that would not necessarily be the same for a given subject if the data-generating process were repeated. In other words, in the traditional view that most political scientists learn, error enters "via repeated drawing of individual error terms," while CLRT conceives of error as entering the model from "repeated random assignments of treatments to subjects" (Kennedy 1995, 86). The CLRT view is not wrong, but it is different, which might not be appealing to applied researchers.[3]

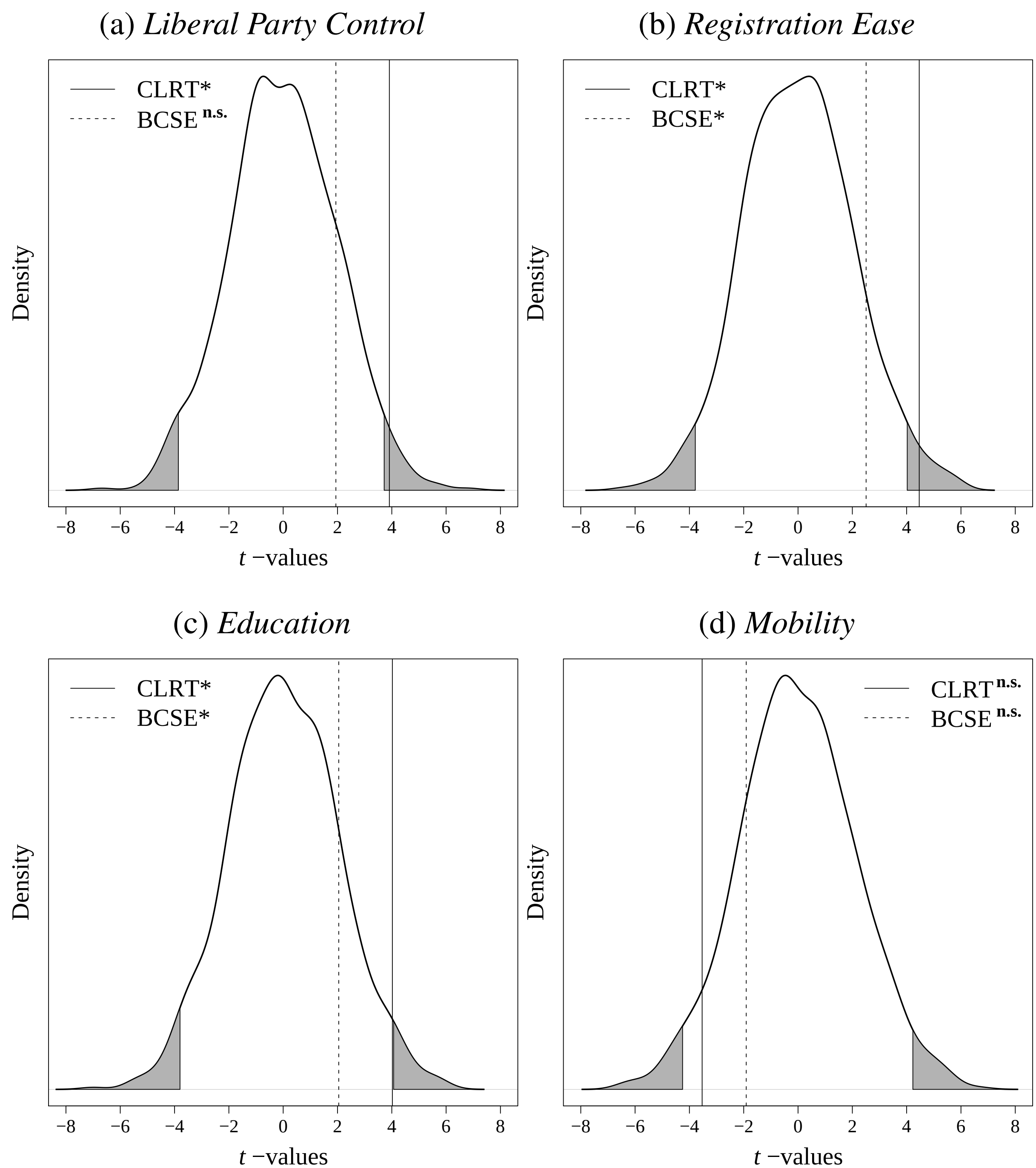
3.3 Ease of Implementation

Finally, the BCSE method is much easier to implement than CLRT. Most statistical packages that perform randomization tests are set up for experimental data because randomization testing is less common with observational data (Kennedy 1995; Kennedy and Cade 1996). In fact, to conduct the simulations and replication reported above, I wrote my own `R` function to perform CLRT. Some researchers may not be interested in doing this for their own work. In contrast, the BCSE method is available with commands in `R` and Stata, as is shown above.

4 Replication with CLRT

I estimated CLRT for the Brown, Jackson, and Wright (1999) model. Results were generally the same between CLRT and BCSE, though there were a few differences. Figure A.1 provides an example with four coefficients. The graphs plot the t-value density from CLRT with 1,000 resamples. The solid and dashed lines represent OLS-SE and BCSE t-values, respectively. The shaded areas represent the 95% rejection region, empirically-derived from the CLRT. The CLRT estimate shows statistical significance if the solid line falls in the shaded region and the BCSE estimate shows significance if the dashed line is larger in absolute value than 1.96.

[3] In addition, recall that CLRT assumes independence and exchangeability of errors between clusters while BCSE only assumes independence. Researchers never know if the exchangeability assumption holds, so a method that does not require it is preferable to one that does.



Note: The graphs plot densities of the t-values from CLRT of the Brown, Jackson, and Wright (1999) model with 1,000 resamples. The solid lines represent OLS-SE t-values and the dashed lines represent BCSE t-values. The CLRT estimate shows statistical significance if the solid line falls in the shaded region and the BCSE estimate shows significance if the dashed line is larger in absolute value than 1.96.

Figure A.1: CLRT and BCSE t-values for Four Coefficients from the Brown, Jackson, and Wright (1999) Model of State Voter Registration

Notice that for *Liberal Party Control*, the main independent variable of interest, the BCSE estimate yields a (slightly) insignificant t-value while CLRT yields a statistically significant result. *Registration Ease* and *Education* are significant with both methods and *Mobility* is not significant with either method. Overall, the two methods are very similar in this case, with BCSE producing a slightly more conservative estimate for the main variable of interest.

References

Brown, Robert D., Robert A. Jackson and Gerald C. Wright. 1999. "Registration, Turnout, and State Party Systems." *Political Research Quarterly* 52(3):463–479.

Erikson, Robert S., Pablo M. Pinto and Kelly T. Rader. 2010. "Randomization Tests and Multi-Level Data in State Politics." *State Politics and Policy Quarterly* 10(2):180–198.

Feng, Ziding, Dale McLerran and James Grizzle. 1996. "A Comparison of Statistical Methods for Clustered Data Analysis with Gaussian Error." *Statistics in Medicine* 15(16):1793–1806.

Fisher, Ronald A. 1922. "On the Interpretation of χ^2 Ffrom Contingency Tables, and the Calculation Of p." *Journal of the Royal Statistical Society* 85(1):87–94.

Kennedy, Peter E. 1995. "Randomization Tests in Econometrics." *Journal of Business & Economic Statistics* 13(1):85–94.

Kennedy, Peter E. and Brian S. Cade. 1996. "Randomization Tests for Multiple Regression." *Communication in Statistics B: Simulation and Computation* 25(4):923–936.

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2):341–355.