Cluster Wild Bootstrapping to Handle Dependent Effect Sizes in Meta-Analysis with a Small Number of Studies

BITSS 10th Annual Meeting

Megha Joshi

2022-02-09

Meta Analysis

- Set of statistical techniques to synthesize results from multiple studies on the same topic
- Goals of meta-analysis
    - Summarize effect size estimates across studies
    - Characterize variability in effect sizes
    - Explain the variability in effect sizes

Dependence

- Typical meta-analytic techniques (like meta-regression) involves the assumption that effect sizes are independent
- However, common for each primary study to yield more than one effect size or studies to be nested in some way creating dependence
- Example: Garrett, Citkowicz, and Williams (2019)
  - Meta-analysis of randomized studies examining the effect of professional learning interventions for teachers on classroom practice
  - Included studies with multiple outcomes measured on the same sample

Handling Dependence

- Ignore dependence
    - Incorrect standard errors, incorrect inference from hypothesis tests
- Ad-hoc methods
    - Selecting one effect per study
    - Analysing subsets of data separately
    - Loss of information
- Standard multivariate methods - ideal
    - Require info on covariance between effect sizes
    - Primary studies often don't report

Robust Variance Estimation

- Cluster robust variance estimation (CR0-type CRVE) (Hedges, Tipton, and Johnson, 2010)
    - Rough approximate assumptions about dependence structure - sandwich estimators
    - Only works well when number of studies is large (> 40, Hedges Tipton, Johnson, 2010)
    - Meta-analysis in social science research typically have smaller number of studies
    - Small number of studies - CRVE - Type 1 error inflation - meta-analysts can conclude some effect is present when it is actually not
    - Example, can conclude teacher professional learning programs are effective for teachers who are older when in reality that effect does not exist

Types of Hypothesis Tests

- Test of single coefficients
  - For example: Does the effect of the professional learning treatment vary depending on the average teaching experience of the samples of teachers in primary studies?
- Multiple-contrast hypothesis tests
  - Do effects differ across different levels of a moderator variable?
  - Garrett, Citkowicz, and Williams (2019) - A multiple-contrast hypothesis could test whether effects of teacher professional learning programs on classroom practice differ for teachers teaching in different grade levels: K-5, 6-8, and 9-12?

Implications of Multiple-Contrast Hypothesis Tests

- If effects differ
  - For example, effects are positive for teachers in elementary and middle schools but null or negative for teachers in high schools
  - Policy implication - re-evaluate the programs for high school teachers and identify ways to make it better or curtail use of program in high schools

Small Sample Corrections

- Tipton (2015) for tests of single coefficients
- Tipton and Pustejovsky (2015) for multiple-contrast hypothesis tests
- Both recommended a method - HTZ test
    - CR2 correction method and using the Satterthwaite degrees of freedom for single coefficient tests
    - Extension of CR2 + Satterthwaite for multiple-contrast hypothesis tests
- HTZ controls Type 1 error rates adequately
- But, really low Type 1 error rates especially for multiple-contrast hypothesis tests (Tipton and Pustejovsky, 2015)
    - Indicating that the test may have low power
    - e.g., might miss the difference in the treatment effect across grade levels when the difference actually exists

Cluster Wild Bootstrapping (CWB)

- Alternative method - examined in the econometrics literature - not in meta-analytic framework
- Bootstrapping - estimate unknown quantities by re-sampling from original data many times (Boos et al., 2013)
- CWB - re-sampling residuals by multiplying them by cluster-level random weights (Cameron, Gelbach, and Miller 2008)

CWB Algorithm

1. Fit a null model and a full model on the original data

2. Obtain residuals from the null model

3. Generate an auxiliary random variable that has mean of 0 and variance of 1 and multiply the residuals by the random variable (e.g., Rademacher weights) set to be constant within clusters (CWB)

   - Can also multiply the residuals by CR2 matrices before multiplying by weights (CWB Adjusted)

4. Obtain new outcome scores by adding the transformed residuals to the predicted values from the null model fit on the original data

5. Re-estimate the full model with the new calculated outcome scores and obtain the test statistic

6. Repeat steps 3-5 R times. Calculate p-value:

$$p = \frac{1}{R}\sum_{r=1}^{R} I\left(F^{(r)} > F\right)$$

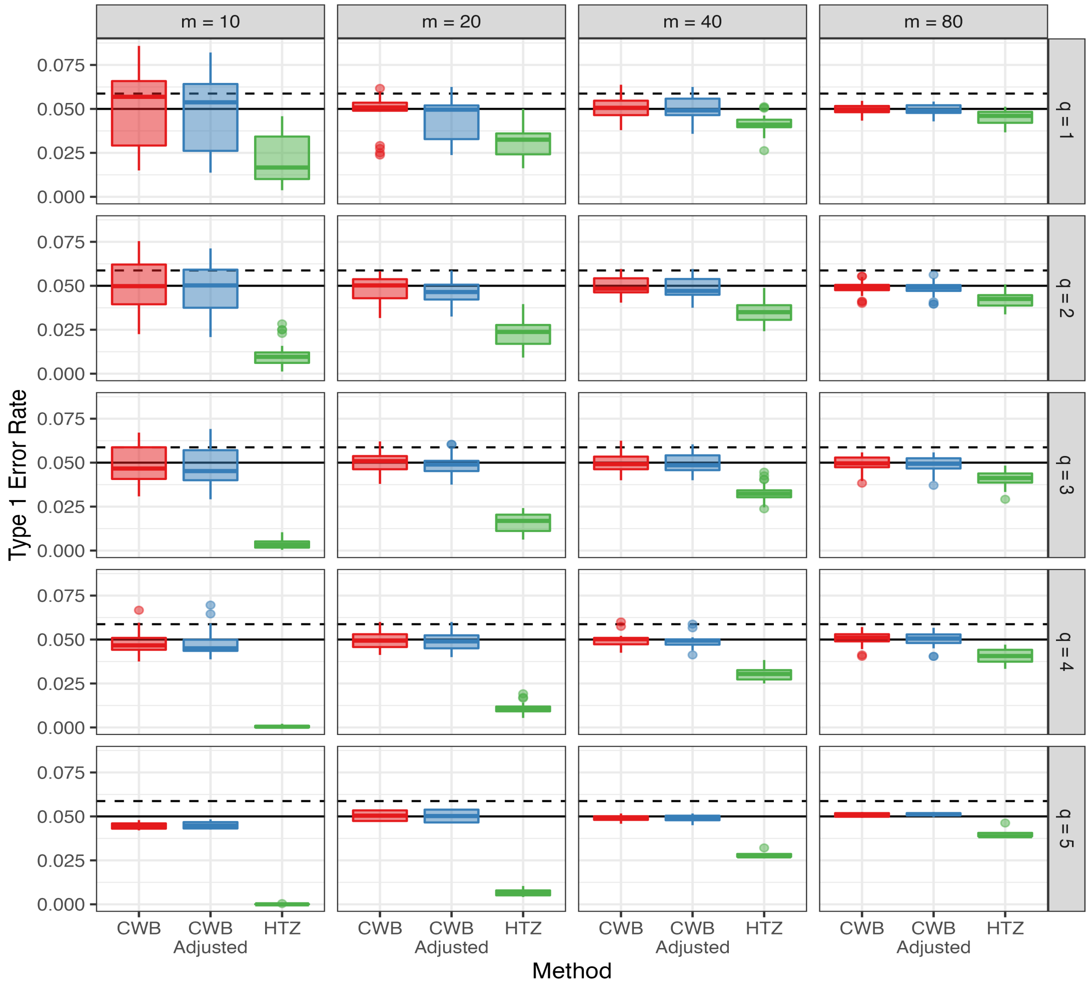
Research Question

To what extent does CWB improve upon the current standard test, the HTZ test, in terms of Type I error rates and power?

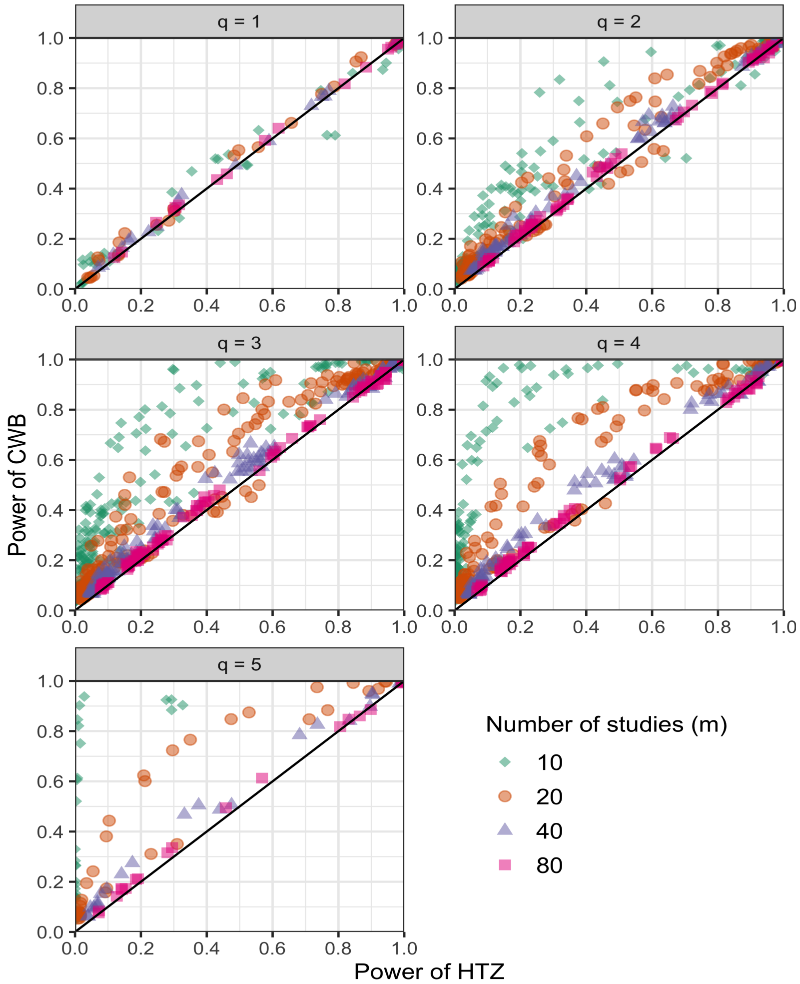
Simulation

- Ran two simulations - presenting on Simulation 1
- Compared CWB against the HTZ test in terms of Type 1 error rates and power
- Results
  - CWB maintained Type 1 error rates adequately
  - And, provided huge gains in power over the standard method, the HTZ test

Results: Type I Error



Results: Relative Power



Conclusion

- Dependent effect sizes - common
- Ignore them - incorrect standard errors and inferences
- Use RVE - Type 1 error inflation - false discovery rate high
- Use small sample correction HTZ test - may miss effects that are present
- Use CWB - balances Type 1 error rates and also provides more power than existing corrections

You can find our preprint here

R Package

https://meghapsimatrix.github.io/wildmeta/

THANK YOU!

Questions?

References

Boos, D. D., & others. (2003). Introduction to the bootstrap world. Statistical Science, 18(2), 168–174.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. The Review of Economics and Statistics, 47.

Garrett, R., Citkowicz, M., & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. Review of research in education, 43(1), 106-137.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods, 1(1), 39–65.

References

Joshi, M., Pustejovsky, J. E., & Beretvas, S. N. (2022). Cluster wild bootstrapping to handle dependent effect sizes in meta-analysis with small number of studies. Research Synthesis Methods. (Accepted).

Tipton, E., & Pustejovsky, J. E. (2015). Small-Sample Adjustments for Tests of Moderators and Model Fit Using Robust Variance Estimation in Meta-Regression. Journal of Educational and Behavioral Statistics, 40 (6), 604–634.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. Psychological Methods, 20(3), 375–393.