Using Cluster Robust Standard Errors (CRSEs) to Analyze Nested Data with a Few Clusters: Extensions

2023.10.13

Francis L. Huang
University of Missouri
huangf@missouri.edu

Bixi Zhang
City University of New York
bzhang2@gc.cuny.edu.

Paper presented at the 23rd International Conference on Educational Research (ICER) Conference
ICER (snu.ac.kr), October 2023, Seoul National University.

Agenda

- Background on data nonindependence and CRSEs
- Research Questions
- Method and Results

Dependent data are common in educational research…

- Also referred to as clustered or nested data
- Violates the regression assumption of observation independence
- Results from:
  - Analyzing clustered data (e.g., students within schools)
  - Analyzing longitudinal data (e.g., outcomes within individuals)

Repercussion of violating observation independence are not negligible…

- This does not get solved by just getting more observations (i.e., increase your sample size)
- Can result in under (also over) estimated standard errors
  - Results in higher Type I or II errors
  - Statistical inference becomes questionable
  - Not a mere technical concern: several studies that have ignored this have issued corrections or differing conclusions after accounting for this

Various approaches have been suggested to account for dependent data…

- Multilevel modeling (mixed effect models)
- Generalized estimating equations (GEEs)
- Using cluster robust/sandwiched standard errors (CRSEs)

CRSEs (Liang & Zeger, 1986) have been commonly used to account for clustering…

- The OLS estimate for the coefficients is:

$$\widehat{\beta} = (X'X)^{-1}(X'y)$$

- The LZ (cluster robust) standard errors can be computed using the diagonal of (with different v$\widehat{\Omega}_g$ions of):

$$var\left(\widehat{\beta}\right) = \Sigma = \left(X'X\right)^{-1} \sum_{g=1}^{G} X_g' \widehat{\Omega}_g X_g \left(X'X\right)^{-1}$$

Huang, F. L., & Li, X. (2022). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods*, *54*, 1181–1199. https://doi.org/10.3758/s13428-021-01627-0

However, it is well known that CRSEs are still underestimated when the number of clusters is low (e.g., < 50)

- At times, the number of clusters is limited:
  - For example: a cluster randomized trial with only a few clusters participating
  - A longitudinal study where there are only a few individuals
- Low cluster situations are common:
  - In a review of 285 CRTs in the health sciences, the median number of clusters was only 21 (Ivers et al., 2021)
  - A review of 78 papers in sociology journals from 2011 to 2014 indicated a quarter of studies had fewer than 20 clusters (Heisig et al., 2017)
- In those cases, there is still the risk of making incorrect statistical inferences when using CRSEs

However, over 20 years ago, Bell and McCaffrey (2002) suggested an adjustment to the LZ CRSE…

- Referred to as the bias reduced linearization approach (BRL)
- Also referred to as the:
  - CR2VE (Cameron & Miller, 2015),
  - CV2 (MacKinnon & Web, 2017),
  - LZ2 (Imbens & Kolesar, 2016),
  - CR2 (Pustejovsky & Tipton, 2018)
- The ‘2’ parallels the HC2 (heteroskedasticity consistent) SEs for sandwiched standard errors which uses the hat (i.e., projection matrix) when computing the ‘meat’ component

Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28,* 169–182.

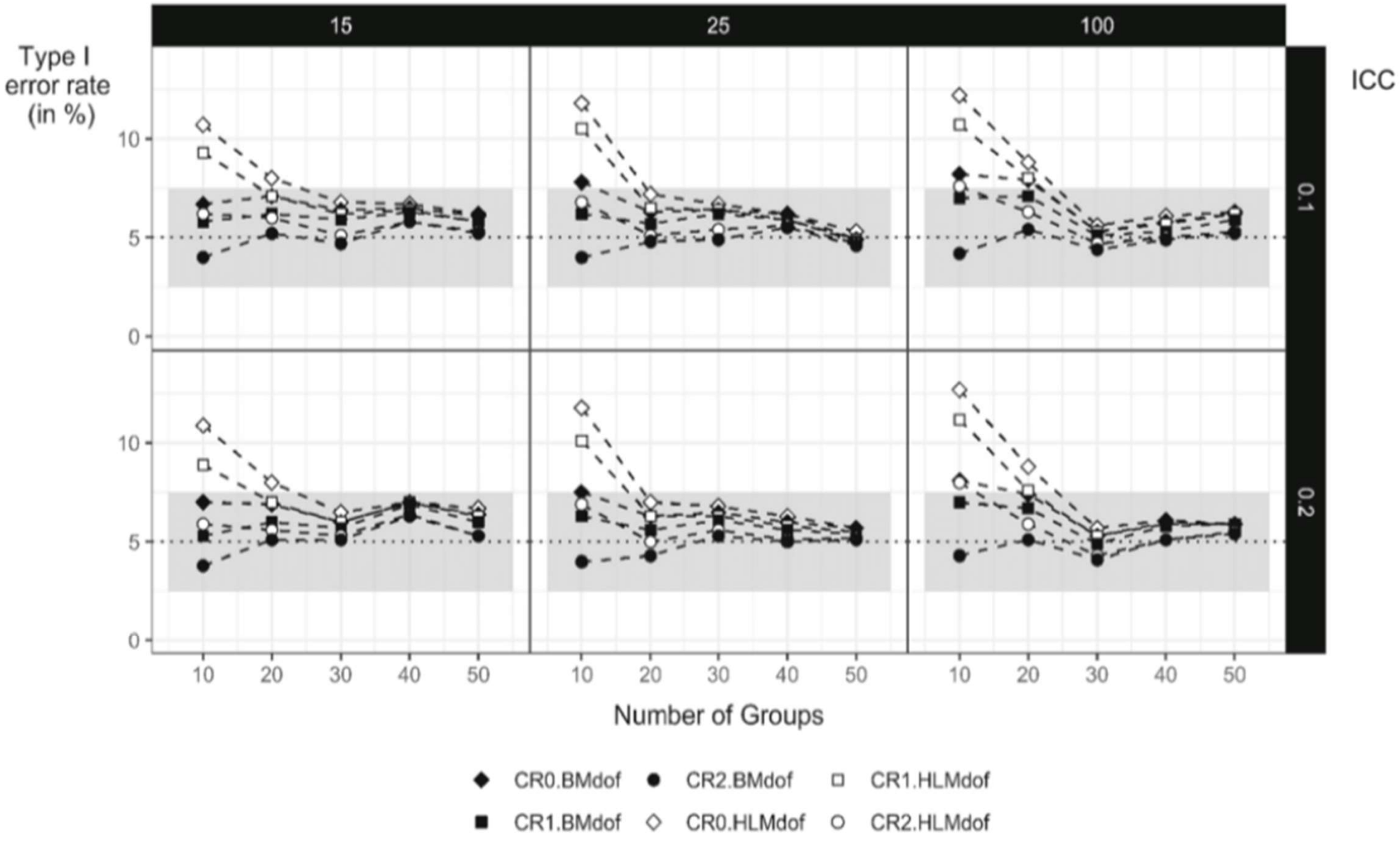
For the CR2, the $\widehat{\Omega}_g$ is…

$$\widehat{\Omega}_g = \left[I_{Ng} - H_{gg}\right]^{-\frac{1}{2}} e_g e_g' \left[I_{Ng} - H_{gg}\right]^{-\frac{1}{2}}$$

- *e* is the residual $(y_i - \hat{y}_i)$
- I_{Ng} is an identify matrix for group g
- H_{gg} is a hat matrix cluster cluster g
- -1/2 refers to the inverse of the symmetric square root
- See Imbens & Kolesar (2016, p. 709)

Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, *98*(4), 701–712.

A prior study has shown that the CR2 (w/dof adjustments) performs well, even with a few clusters (here compared with a mixed model) (see Huang & Li, 2022 for details)



We extended the ability to compute the CR2 (and dof adjustments) using SPSS using a GUI

- Despite being very popular, SPSS is limited with the CRSEs that it can compute
- Can download our free add on from:

https://github.com/flh3/CR2

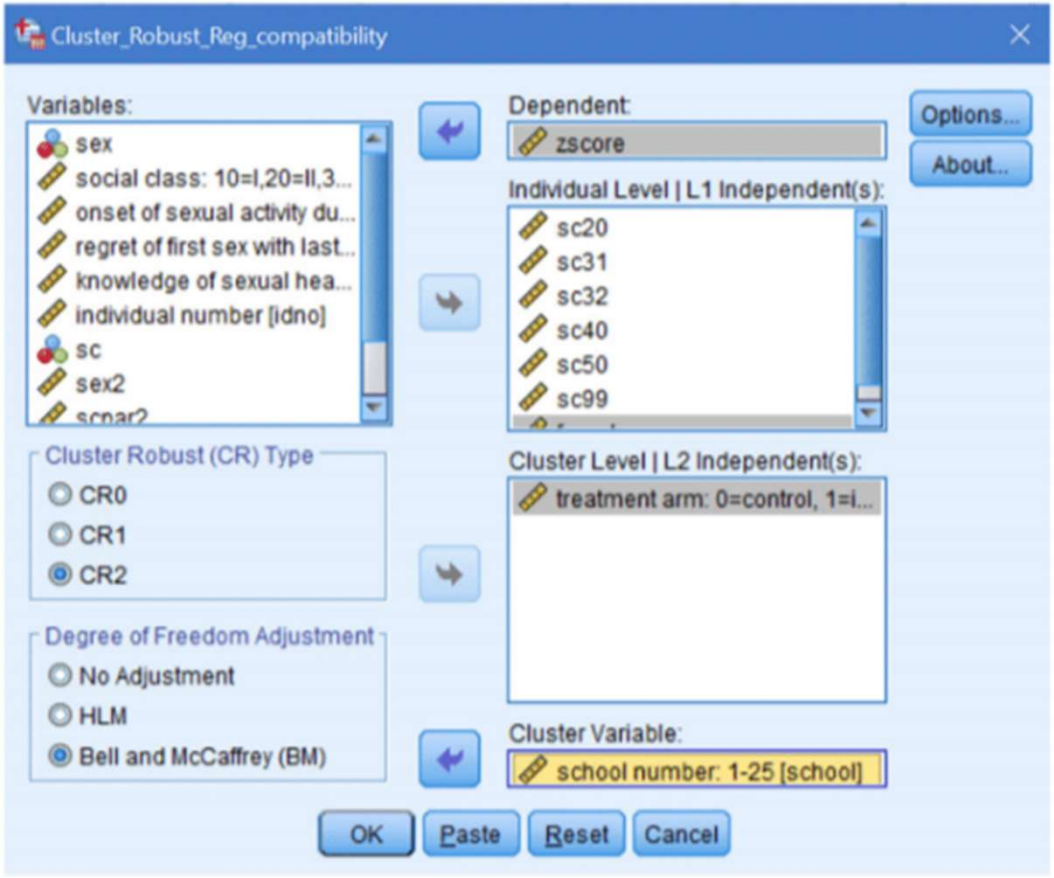
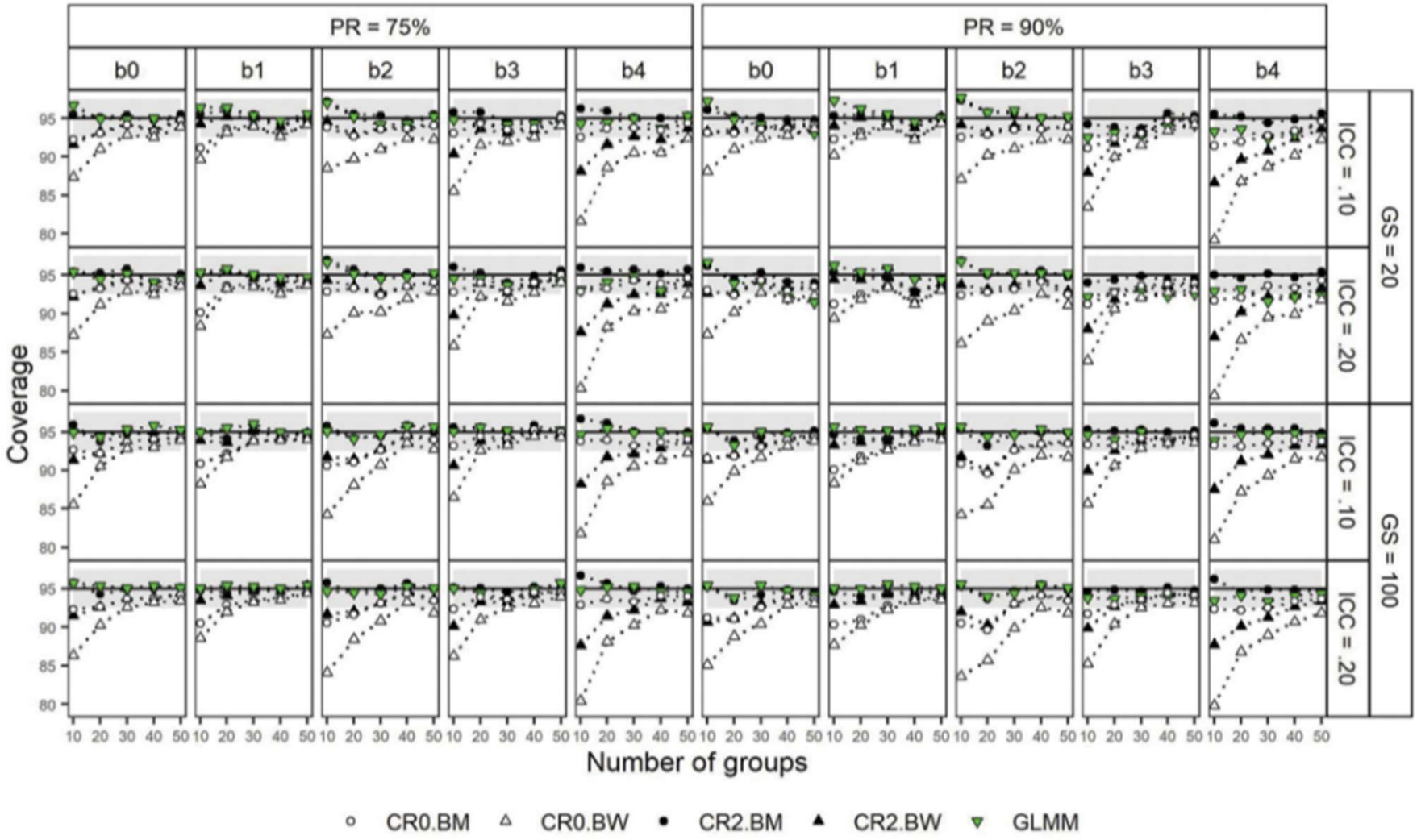


Fig. 6 Custom Cluster-Robust Regression Analysis dialog box in SPSS

The CR2 also works well with binary outcomes (i.e., logistic regression models)



- Huang, F., Zhang, B., & Li, X. (2023). Using robust standard errors for the analysis of binary outcomes with a small number of clusters. *Journal of Research on Educational Effectiveness*, *16*(2), 213-245. https://doi.org/10.1080/19345747.2022.2100301

We also extended our CR2 add on for SPSS to work with binary outcomes…

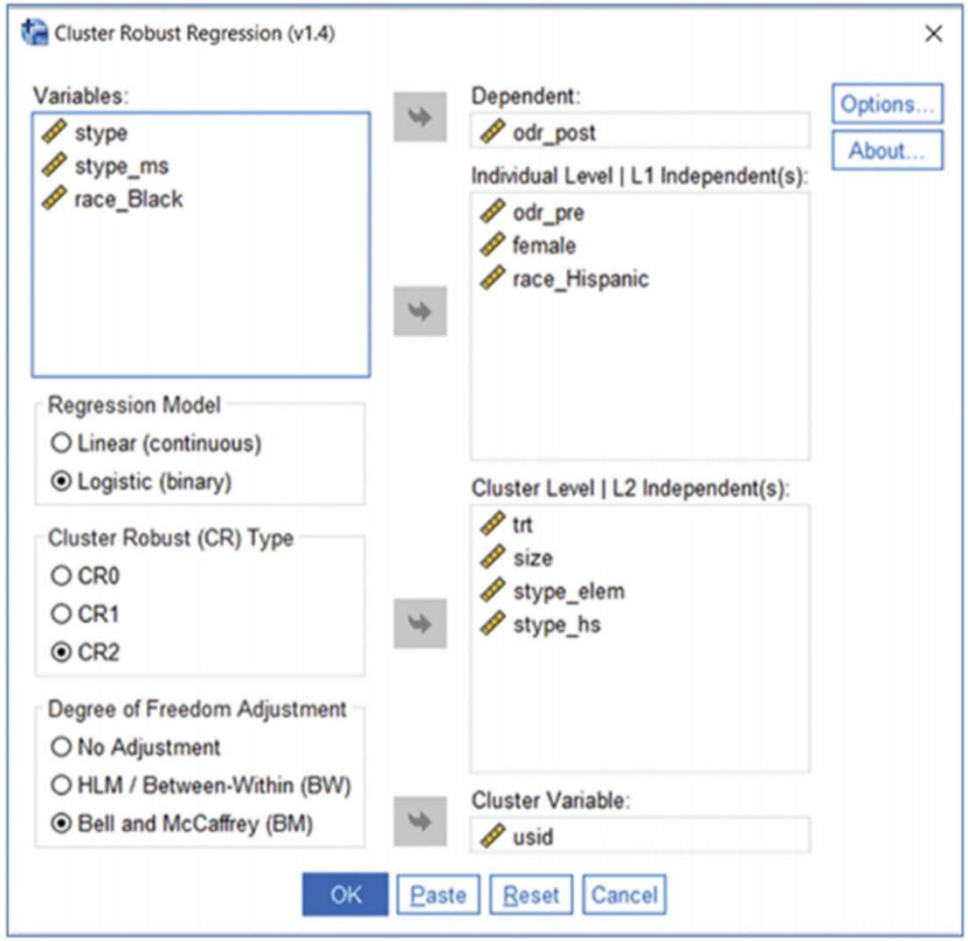


Figure 4. SPSS interphase for specifying a logistic regression model using cluster robust standard errors.

- Huang, F., Zhang, B., & Li, X. (2023). Using robust standard errors for the analysis of binary outcomes with a small number of clusters. *Journal of Research on Educational Effectiveness*, *16*(2), 213-245. https://doi.org/10.1080/19345747.2022.2100301

Our work has also been extended to multilevel models

- Q: Aren't mixed models SEs already robust?
  - Yes: to account for nonindependence
  - No: if there is heteroskedasticity
- The advantage of CRSEs is that they are robust to both nonindependence AND heteroskedasticity
- However, CRSEs (vs model-based SEs) in a mixed model also have issues with underestimation when the number of clusters is low (CR0)

Huang, F., Wiedermann, W., & Zhang, B. (2023). Accounting for heteroskedasticity resulting from between-group differences in multilevel models. *Multivariate Behavioral Research, 58,* 637-657. doi: 10.1080/00273171.2022.2077290.

For CRSEs for MLMs…

- MLM can be written as:

$$\boldsymbol{y_c} = \boldsymbol{X_c\beta} + \boldsymbol{Z_c u_c} + \boldsymbol{\varepsilon_c}$$

- Using ML/REML, the model-based variance of the outcome for cluster c is:

$$\boldsymbol{V}_c = \widehat{Var}(\boldsymbol{y}_c) = \boldsymbol{Z}_c\hat{\boldsymbol{T}}\boldsymbol{Z}_c^{'} + \hat{\boldsymbol{R}}_c$$

- And the estimate of the FE is:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}.$$

For CRSEs for MLMs (cont.)…

- As proposed by Liang & Zeger (1986):

$$\widehat{Var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\sum_{c=1}^{C}\boldsymbol{X}_c'\boldsymbol{V}_c^{-1}\hat{\boldsymbol{\Omega}}_c\boldsymbol{V}_c^{-1}\boldsymbol{X}_c(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}.$$

- The $\hat{\boldsymbol{\Omega}}_c$ uses the residuals per cluster (squared residuals on the diagonal)

with the CR2 estimator. However, to extend this to a mixed model framework (McCaffrey et al., 2001; Pustejovsky & Tipton, 2018; Tipton, 2015), the inverse of the $\boldsymbol{V}_c$ matrix should also be incorporated as well:

$$\sum_{c=1}^{C}\boldsymbol{X}_c'\boldsymbol{V}_c^{-1}[\boldsymbol{I}_c - \boldsymbol{H}_c]^{-\frac{1}{2}}\hat{\boldsymbol{e}}_c\hat{\boldsymbol{e}}_c'[\boldsymbol{I}_c - \boldsymbol{H}_c]^{-\frac{1}{2}}\boldsymbol{V}_c^{-1}\boldsymbol{X}_c \qquad (10)$$

where the hat matrix can also be specified as

$$\boldsymbol{H}_c = \boldsymbol{X}_c(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}_c'\boldsymbol{V}_c^{-1} \qquad (11)$$

(Tipton, 2015).

Agenda

- Background on data nonindependence and CRSEs
- Research Questions
- Method and Results

Other questions– can the CR2 be used with …

Other types of nested data

- Study 1: Cross classified (CC) models? (two-way nesting)
- Study 2: Longitudinal data?

Agenda

- Background on data nonindependence and CRSEs
- Research Questions
- Method and Results

1) For cross classified (CC) data– there can be two (or more levels) of clustering at the same level…

- Students nested within Math and English teachers[1]
- Students nested within neighborhoods and schools[2]
- Students nested within primary and secondary schools[3]

[1]Gregory, A., & Huang, F. (2013). It takes a village: The effects of 10th grade college-going expectations of students, parents, and teachers four years later. *American Journal of Community Psychology*, *52*(1–2), 41–55. https://doi.org/10.1007/s10464-013-9575-5

[2]Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*, *18*(4), 321–349. https://doi.org/10.2307/1165158

[3]Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, *15*(4), 337–350.

As there are two forms of clustering, one set of clusters may have a low number…

- Rasbash and Goldstein (1994): units within 148 primary schools but only 19 secondary schools
- Carroll-Scott et al. (2015): units within 12 schools and 25 neighborhoods
- Claus et al. (2020): scores within 127 participants but only 12 raters

- Thus– using TW-CR2 should be robust if used with a few clusters as well…

The modification to the CRSE is straightforward as presented by Cameron et al. (2011) and Thompson (2011)

- The standard CR0 correction is just the robust SE for cluster K + the robust SE for cluster J – the intersection of the robust SE for cluster JK

$$\hat{V}(\hat{\beta})_{TW-CR0} = (X'X)^{-1}X'(\hat{\varepsilon}\hat{\varepsilon}' \cdot S^J)X\,(X'X)^{-1} + (X'X)^{-1}X'(\hat{\varepsilon}\hat{\varepsilon}' \cdot S^K)X\,(X'X)^{-1}$$
$$- (X'X)^{-1}X'(\hat{\varepsilon}\hat{\varepsilon}' \cdot S^{J\cap K})X\,(X'X)^{-1}.$$

- For the CR2, the CRSEs are computed as shown earlier (using the hat matrices)

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway Clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249. https://doi.org/10.1198/jbes.2010.07136

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, *99*(1), 1–10. https://doi.org/10.1016/j.jfineco.2010.08.016

Using a simulation where the number of cluster J and K were altered and type of CC was tested (complete or partial)…

- Results showed that the TW-CR2 could be effective, but in conditions where 50-20 (or 30-40) clusters were present
  - Undercoverage still possible with 50-10 (or 30-30) for example
- CCREM performed well even with a limited number of clusters

- Zhang, B., & Huang, F. (in press). Investigating the use of robust standard errors to account for two-way clustering in cross-classified data structures. *Dependent Data Analysis, 2nd edition*. Springer.

2) For longitudinal data, when using multilevel models to analyze data, often…

- A composite model (explain) may be written:

$$Y_{ti} = \gamma_{00} + \gamma_{01} trt_i + \gamma_{10} time_{ti} + \gamma_{11} trt_i \times time_{ti} + u_{01} + u_{1i} time_{ti} + r_{ti}$$

- Slopes may be fixed or allowed to randomly vary (random slope models)
- The R (residual covariance) matrix may be assumed to be homoscedastic or have some other structure (e.g., autoregressive, toeplitz)

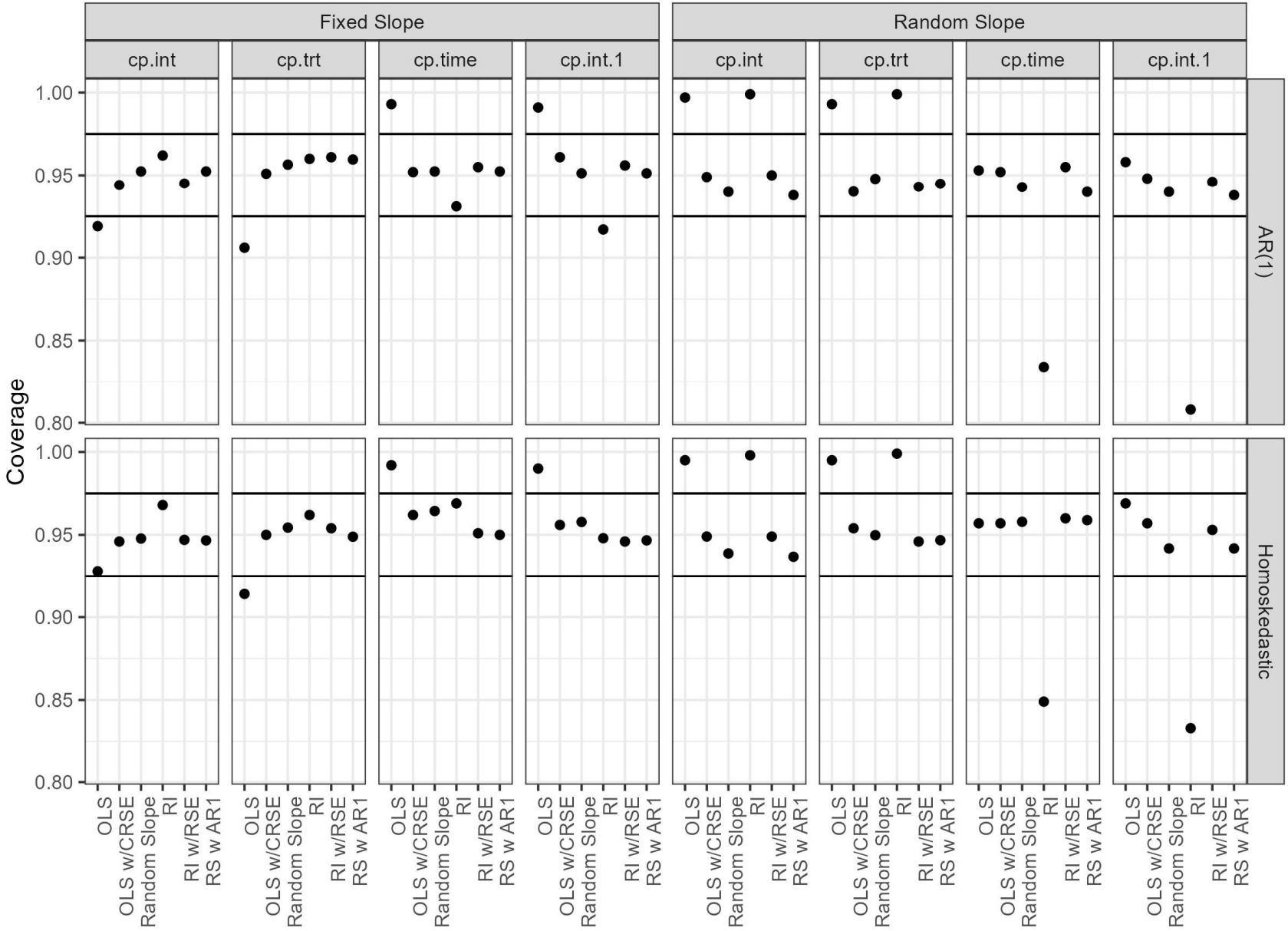
Conducted a simulation where…

Data were generated where:

- Slopes were fixed or allowed to vary (random)
- R matrix was specified to be homoscedastic or autoregressive (AR1)
- # of clusters varied (n = 30, 50, 100)
- # of time points varied (4 or 8)

- Analyzed using:
  - OLS
  - OLS with CR2
  - RI
  - RI with CR2
  - RS
  - RS with AR1 specified

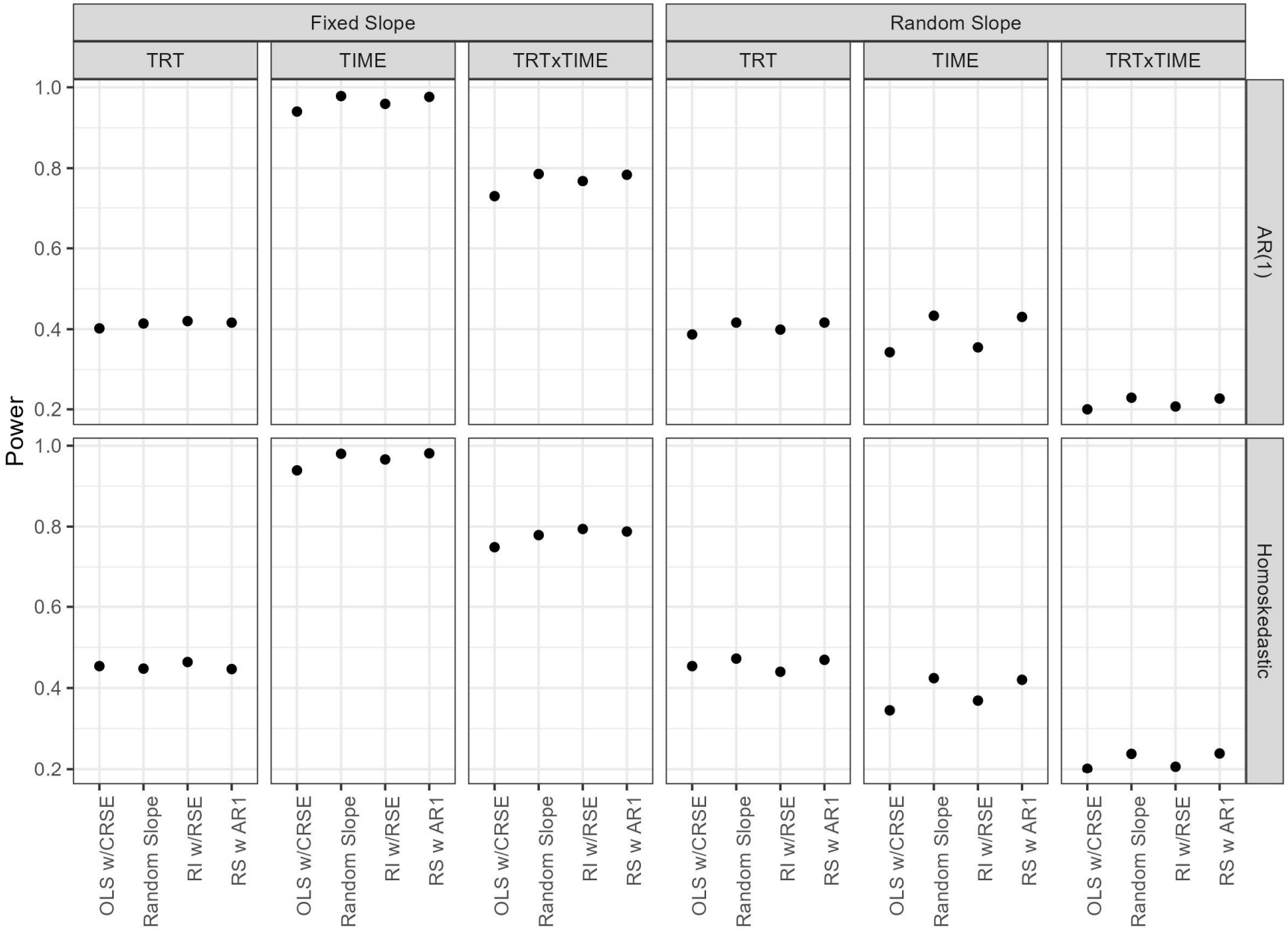
Preliminary results (n = 30)…robust SEs are helpful! (Talk through…)



Takeaways

- OLS here is not really a focus– helpful to include to see if the sim is working as it should be (L2 is underestimated, L1 are overestimated) ✓
- RI doesn't work well when there is a random slope (but we knew that already) ✓
- CRSEs work well– even when used with OLS and a simple RI model
- The R matrix modification (AR1 vs homoscedastic) does not seem to affect results much– need to investigate a bit more…

For acceptable methods, power is comparable (n = 30) (Talk through…)



Overall (for longitudinal studies)…

- CRSEs provide a simple approach to addressing certain questions when it comes to fitting growth models (e.g., including random slopes, worrying about R matrix)
  - Avoid the need to test competing models to find best fitting model
  - Can be used as a diagnostic as well…
- Can be used with basic OLS models and with RI models
- CRSEs are slightly less powerful for L1 effects compared to RS models

Using Cluster Robust Standard Errors to Analyze Nested Data with a Few Clusters: Extensions

2023.10.13

Francis L. Huang
University of Missouri
huangf@missouri.edu

Bixi Zhang
City University of New York
bzhang2@gc.cuny.edu.

College of Education & Human Development
University of Missouri

THE CITY UNIVERSITY OF NEW YORK

Paper presented at the 23rd International Conference on Educational Research (ICER) Conference
ICER (snu.ac.kr), October 2023, Seoul National University.