

Small-sample cluster-robust variance estimators for two-stage least squares models

James E. Pustejovsky
The University of Texas at Austin

Context

In randomized field trials of educational interventions, participants do not always comply with their assigned treatment condition. Under these circumstances, researchers might adopt an intent-to-treat (ITT) perspective and focus on the average effect of being *assigned* to intervention. Alternately, researchers might focus on the complier average causal effect (CACE), which is the average effect of *receiving* the intervention for the subgroup of participants for whom treatment assignment induces treatment receipt (Angrist, Imbens, & Rubin, 1996). There is an inherent interpretive trade-off between these two perspectives: whereas ITT analysis has strong internal validity, CACE analysis may have clearer construct validity because it relates directly to treatment receipt. Two-stage least squares (2SLS) estimation is a standard approach for CACE analysis—indeed, it is the *only* estimation method accepted in the most recent What Works Clearinghouse Standards (What Works Clearinghouse, 2018).

Inference in ITT or CACE analysis is often based on cluster-robust variance estimation (CRVE) (Cameron & Miller, 2015). CRVE is used for analyzing cluster-randomized trials in order to account for within-cluster dependence. CRVE may also be used in analysis of multi-site trials (clustering by site) when the goal is to generalize to a larger population of sites (Abadie, Athey, Imbens, & Wooldridge, 2017). In both settings, CRVE provides a way to avoid imposing strong distributional assumptions about the error structure in ITT or CACE estimating equations.

Conventional CRVE methods require a large number of clusters to obtain properly calibrated hypothesis tests and confidence intervals. Further, rejection/coverage rates from CRVE depend on features of the study design beyond just the total number of clusters (Bell & McCaffrey, 2002; Pustejovsky & Tipton, 2016). Bell and McCaffrey introduced modifications of CRVE, known as bias-reduced linearization estimators, which have improved rejection/coverage rates even with a small number of clusters. Their initial work was limited to ordinary and weighted least squares regression (Bell & McCaffrey, 2002; McCaffrey, Bell, & Botts, 2001), but the methods have subsequently been extended to generalized estimating equations (McCaffrey & Bell, 2006) and fixed effects regression (Pustejovsky & Tipton, 2016).

Aims

Small-sample CRVE methods can readily be applied to ITT analysis in multi-site and cluster-randomized trials. However, similar methods are not available for CACE analysis. Researchers

aiming to estimate CACE thus face another trade-off: should they use conventional CRVE, despite the possibility of obtaining mis-calibrated tests and confidence intervals, or turn to methods that impose stronger distributional assumptions? To help resolve this conundrum, we propose bias-reduced linearization estimators for 2SLS models and evaluate the type-I error rates of the methods in Monte Carlo simulations.

Methods

Let us first consider a linear regression model for data from a sample of J clusters, where cluster j includes n_j individual observations:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{e}_j \quad (1)$$

For simplicity, we focus on the ordinary least squares estimator, $\hat{\boldsymbol{\beta}} = \mathbf{M} \sum_{j=1}^J \mathbf{X}'_j \mathbf{y}_j$, where $\mathbf{M}_\mathbf{X} = (\sum_{j=1}^J \mathbf{X}'_j \mathbf{X}_j)^{-1}$. Let $\hat{\mathbf{e}}_j = \mathbf{y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}}$. The bias-reduced linearization estimator of $\text{Var}(\hat{\boldsymbol{\beta}})$ is given by

$$\mathbf{V}^{OLS} = \mathbf{M}_\mathbf{X} \left(\sum_{j=1}^J \mathbf{X}'_j \mathbf{A}_j \hat{\mathbf{e}}_j \hat{\mathbf{e}}'_j \mathbf{A}_j \mathbf{X}_j \right) \mathbf{M}_\mathbf{X}. \quad (2)$$

The \mathbf{A}_j are symmetric matrices defined such that $E(\mathbf{V}^{OLS}) = \text{Var}(\hat{\boldsymbol{\beta}})$ under an analyst-specified working model for $\text{Var}(\mathbf{e}_j)$, $j = 1, \dots, J$. For instance, the analyst might use a “working independence” model with $\text{Var}(\mathbf{e}_j) = \sigma^2 \mathbf{I}_{n_j}$; Imbens and Kolesar (2015) propose some alternative working models. Inference on the regression coefficient β_k is based on the statistic $\beta_k / \sqrt{V_{kk}^{OLS}}$ with a $t(\eta)$ reference distribution, where $\eta = 2E^2(V_{kk}^{OLS})/V(V_{kk}^{OLS})$ is a Satterthwaite-type degrees of freedom approximation that is derived under the working model for $\text{Var}(\boldsymbol{\epsilon}_j)$.

Let us now consider the model

$$\begin{aligned} \mathbf{y}_j &= \mathbf{Z}_j \boldsymbol{\delta} + \mathbf{u}_j \\ \mathbf{Z}_j &= \mathbf{X}_j \boldsymbol{\gamma} + \mathbf{v}_j \end{aligned} \quad (3)$$

where \mathbf{Z}_j includes the endogenous regressor (i.e., the compliance indicator) and additional covariates and \mathbf{X}_j includes the instrument (i.e., treatment assignment indicator) and additional covariates. The 2SLS estimator of $\boldsymbol{\delta}$ is

$$\hat{\boldsymbol{\delta}} = \mathbf{M}_\mathbf{Z} \left(\sum_{j=1}^J \tilde{\mathbf{Z}}'_j \mathbf{y}_j \right), \quad (4)$$

where $\mathbf{M}_Z = (\sum_{j=1}^J \tilde{\mathbf{Z}}_j' \tilde{\mathbf{Z}}_j)^{-1}$ and $\tilde{\mathbf{Z}}_j = \mathbf{X}_j (\sum_{j=1}^J \mathbf{X}_j' \mathbf{X}_j)^{-1} (\sum_{j=1}^J \mathbf{X}_j' \mathbf{Z}_j)$. Let $\hat{\mathbf{u}}_j = \mathbf{y}_j - \mathbf{Z}_j \hat{\boldsymbol{\delta}}$. We propose a bias-reduced linearization estimator of $\text{Var}(\hat{\boldsymbol{\delta}})$ as

$$\mathbf{V}^{2SLS} = \mathbf{M}_Z \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{A}_j \hat{\mathbf{u}}_j \hat{\mathbf{u}}_j' \mathbf{A}_j \mathbf{Z}_j \right) \mathbf{M}_Z, \quad (5)$$

where the adjustment matrices (and Satterthwaite degrees of freedom for hypothesis tests) are calculated as for an ordinary linear regression model of the second stage, $\mathbf{y}_j = \tilde{\mathbf{Z}}_j \boldsymbol{\delta} + \tilde{\mathbf{u}}_j$, under a working model for $\text{Var}(\tilde{\mathbf{u}}_j), j = 1, \dots, J$. For models with a single instrument, the proposed adjustment yields an exactly unbiased estimator of the delta-method approximation to the Wald representation of the 2SLS estimator.

Simulation Design and Results

We designed simulations to evaluate the performance of the proposed method compared to the conventional CRVE for 2SLS. The conventional CRVE is calculated as in (5) but omits the adjustment matrices and uses a standard normal reference distribution. Initial simulations emulated a cluster-randomized trial with cluster-level non-compliance that was correlated with cluster-level mean outcomes to varying degrees, such that naïve OLS estimation yields biased estimates of the CACE. We also varied the total number of clusters in the trial, the fraction of clusters allocated to treatment, and the overall compliance rate. The intra-class correlation of the outcome was set to 0.2 throughout, and cluster sizes were simulated as $n_j \sim 1 + \text{Pois}(10)$. We estimate the CACE using unweighted 2SLS and compare the Type I error rates of hypothesis tests based on the conventional CRVE and the proposed bias-reduced linearization estimator.

Figures 1 and 2 depict the Type I error rates of tests for the CACE estimated by 2SLS using either the conventional or modified test, at α -levels of .05 and .01, respectively. Each line represents a different degree of compliance–outcome correlation. At both α -levels, the conventional test has excessive Type-I error when the total number of clusters is small or treatment allocation is imbalanced—particularly for higher compliance rates. In contrast, the modified method yields error rates at or below the nominal level, with as few as 6 treated and 14 control clusters. The tests become more conservative when compliance levels are lower.

Conclusions

These simulation results provide initial evidence that the proposed small-sample adjustments improve Type I error rates of hypothesis tests for the CACE in the context of cluster-randomized trials with cluster-level non-compliance. In ongoing work, we are examining the performance of these methods in individually-randomized multi-site trials, with individual-level non-compliance that varies across sites. In this context, we will examine performance of the conventional and

modified tests for 2SLS with a single instrument (treatment assignment) and instrument-by-site interactions. We expect that all methods will perform poorly in the latter case due to the inconsistency of 2SLS as the number of instruments increases (Bekker, 1994).

References

- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When Should You Adjust Standard Errors for Clustering?* <https://doi.org/10.3386/w24003>
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. Retrieved from <http://www.jstor.org/stable/2291629>
- Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, *62*(3), 657–681. <https://doi.org/10.2307/2951662>
- Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*(2), 169–181. Retrieved from <http://www.statcan.gc.ca/pub/12-001-x/2002002/article/9058-eng.pdf>
- Cameron, R. C., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, *50*(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Imbens, G. W., & Kolesar, M. (2015). *Robust standard errors in small samples: some practical advice*.
- McCaffrey, D. F., & Bell, R. M. (2006). Improved hypothesis testing for coefficients in generalized estimating equations with small samples of clusters. *Statistics in Medicine*, *25*(23), 4081–98. <https://doi.org/10.1002/sim.2502>
- McCaffrey, D. F., Bell, R. M., & Botts, C. H. (2001). Generalizations of biased reduced linearization. In *Proceedings of the Annual Meeting of the American Statistical Association*.
- Pustejovsky, J. E., & Tipton, E. (2016). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 1–12. <https://doi.org/10.1080/07350015.2016.1247004>
- What Works Clearinghouse. (2018). *What Works Clearinghouse Standards Handbook, Version 4*. Washington, DC.

Figure 1
Type I error rates of tests based on conventional and modified cluster-robust variance estimators at $\alpha = .05$

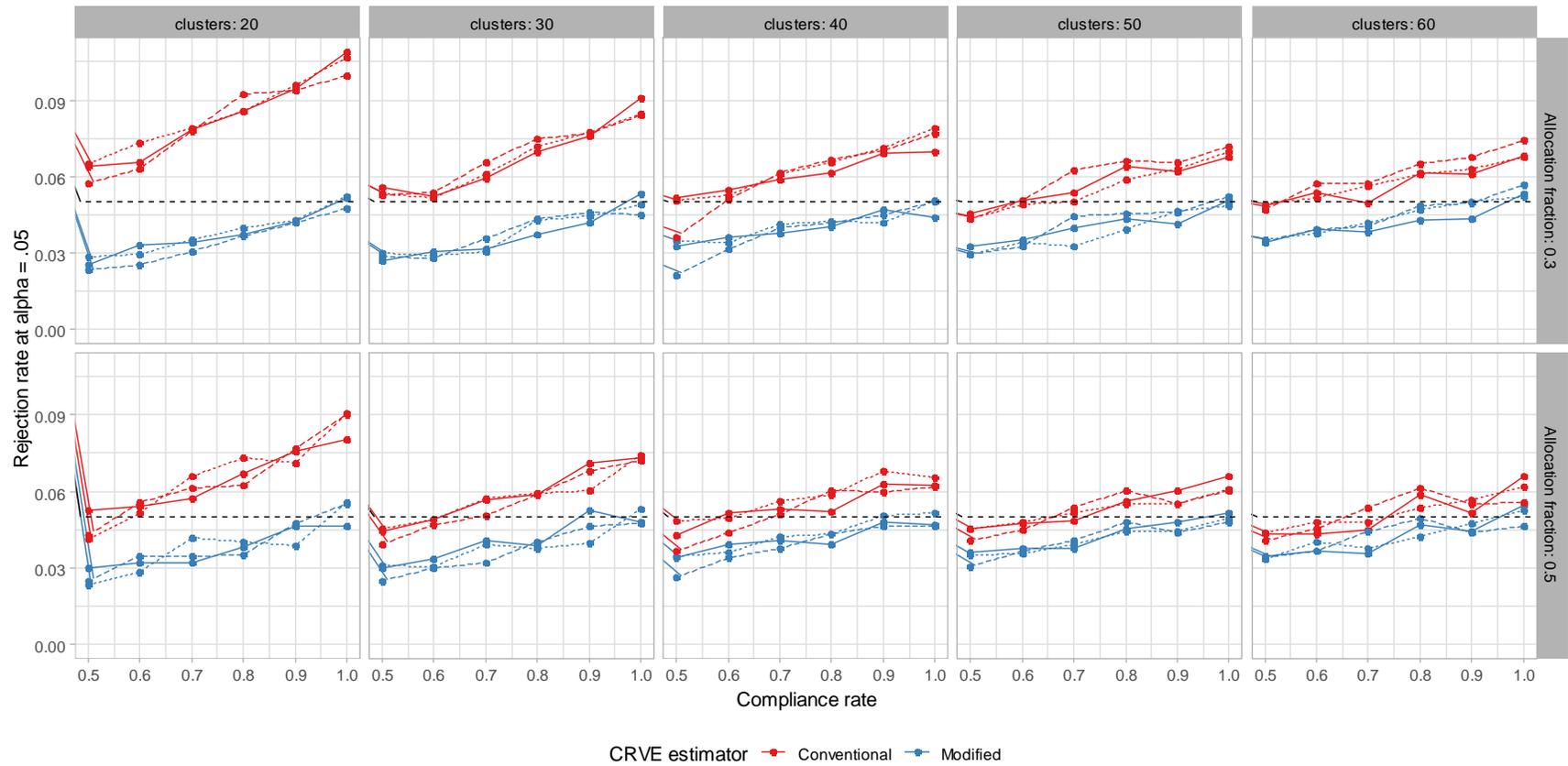


Figure 2

Type I error rates of tests based on conventional and modified cluster-robust variance estimators at $\alpha = .01$

