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

James E. Pustejovsky

March 8, 2019

Randomized trials with non-compliance

- Randomized field trials often encounter **non-compliance** with treatment assignments.
- An initial tension:
  - **Intent-to-Treat** analysis for the average effects of ***treatment assignment***
  - **Instrumental variables** analysis for the ***complier average treatment effect (CATE)***
- **Two-Stage Least Squares** is standard approach for estimating CATE.

Cluster-robust variance estimation (CRVE)

- Common approach to obtaining standard errors/hypothesis tests/confidence intervals for impact estimates.
- Account for dependence without imposing distributional assumptions.
    - Within-cluster dependence in cluster-randomized trials.
    - Site-level heterogeneity in multi-site trials (Abadie, Athey, Imbens, & Wooldridge, 2017).
- Conventional CRVE requires a large number of clusters.
- **Bias-reduced linearization** CRVE methods (Bell and McCaffrey, 2002) work well in small samples.
    - Weighted least squares linear regression (McCaffrey, Bell, & Botts, 2001)
    - Generalized estimating equations (McCaffrey & Bell, 2006)
    - Linear fixed effects models (Pustejovsky & Tipton, 2016)
    - But not for 2SLS

Aim

Develop bias-reduced linearization estimators for 2SLS estimators.

Outline

- Review bias-reduced linearization for OLS models
- Explain approach for 2SLS
- Some simulation results

Ordinary least squares

A linear regression model for data from J clusters:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{e}_j$$

where $\text{Var}\left(\mathbf{e}_j\right) = ???$

The OLS estimator:

$$\hat{\boldsymbol{\beta}} = \mathbf{B_X}\sum_j \mathbf{X}_j'\mathbf{y}_j \qquad \text{where} \qquad \mathbf{B_X} = \left(\sum_j \mathbf{X}_j'\mathbf{X}_j\right)^{-1}$$

Conventional CRVE (sandwich estimator) of $\text{Var}(\hat{\boldsymbol{\beta}})$:

$$\mathbf{V}^{CR0} = \mathbf{B_X}\left(\sum_j \mathbf{X}_j'\hat{\mathbf{e}}_j\hat{\mathbf{e}}_j'\mathbf{X}_j\right)\mathbf{B_X}$$

Bias-reduced linearization

1. Make a "working" assumption that $\text{Var}(\mathbf{e}_j) = \mathbf{\Omega}_j$ for $j = 1, ..., J$.

2. Add extra fillings to the sandwich estimator:

$$\mathbf{V}^{CR2} = \mathbf{B_X} \left(\sum_j \mathbf{X}_j' \mathbf{A}_j \hat{\mathbf{e}}_j \hat{\mathbf{e}}_j' \mathbf{A}_j' \mathbf{X}_j \right) \mathbf{B_X}$$

   where $\mathbf{A}_j$ are chosen so that

$$\text{E}\left(\mathbf{V}^{CR2}\right) = \text{Var}(\hat{\boldsymbol\beta})$$

   under the working model.

- It turns out that this works ***even when the working model is misspecified***.

Two-stage least squares

The model for cluster $j = 1, \ldots, J$:

$$\mathbf{y}_j = \mathbf{Z}_j\boldsymbol{\delta} + \mathbf{u}_j$$
$$\mathbf{Z}_j = \mathbf{X}_j\boldsymbol{\gamma} + \mathbf{v}_j$$

where

- $\mathbf{Z}_j$ includes endogenous regressor (compliance indicator)
- $\mathbf{X}_j$ includes the instrument (treatment assignment)

Two-stage least squares estimation

- **First stage** (appetizer):

$$\mathbf{Z}_j = \mathbf{X}_j \boldsymbol{\gamma} + \mathbf{v}_j$$

with fitted values

$$\tilde{\mathbf{Z}}_j = \mathbf{X}_j \hat{\boldsymbol{\gamma}} = \mathbf{X}_j \mathbf{B}_{\mathbf{X}} \sum_j \mathbf{X}_j' \mathbf{Z}_j$$

- **Second stage** (main course):

$$\mathbf{y}_j = \tilde{\mathbf{Z}}_j \boldsymbol{\delta} + \tilde{\mathbf{u}}_j$$

estimated as

$$\hat{\boldsymbol{\delta}} = \mathbf{B}_{\mathbf{Z}} \sum_j \tilde{\mathbf{Z}}_j' \mathbf{y}_j \qquad \text{where} \qquad \mathbf{B}_{\mathbf{Z}} = \left(\sum_j \tilde{\mathbf{Z}}_j' \tilde{\mathbf{Z}}_j \right)^{-1}$$

Bias-reduced linearization for 2SLS

- CRVE with adjustment matrices:

$$\mathbf{V}^{CR2} = \mathbf{B_Z} \left(\sum_j \tilde{\mathbf{Z}}_j' \mathbf{A}_j \hat{\mathbf{u}}_j \hat{\mathbf{u}}_j' \mathbf{A}_j' \tilde{\mathbf{Z}}_j \right) \mathbf{B_Z}$$

  where $\hat{\mathbf{u}}_j = \mathbf{y}_j - \mathbf{Z}_j \hat{\boldsymbol{\delta}}$.

- Proposal: calculate adjustment matrices $\mathbf{A}_j$ ***based on the second stage only***, for

$$\mathbf{y}_j = \tilde{\mathbf{Z}}_j \boldsymbol{\delta} + \tilde{\mathbf{u}}_j,$$

  under a working model for $\tilde{\mathbf{u}}_j$.

Single instrument IV

With a single-dimensional instrument, CATE is a ratio:

$$\delta = \frac{\beta}{\gamma} = \frac{\text{ITT effect}}{\text{Compliance effect}} \qquad \text{and} \qquad \hat{\delta} = \frac{\hat{\beta}}{\hat{\gamma}}$$

Delta-method approximation to $\text{Var}(\hat{\delta})$:

$$\text{Var}(\hat{\delta}) \approx \frac{1}{\gamma^2}\left[\text{Var}(\hat{\beta}) + \delta^2 \text{Var}(\hat{\gamma}) - 2\delta \text{Cov}(\hat{\beta}, \hat{\gamma})\right]$$

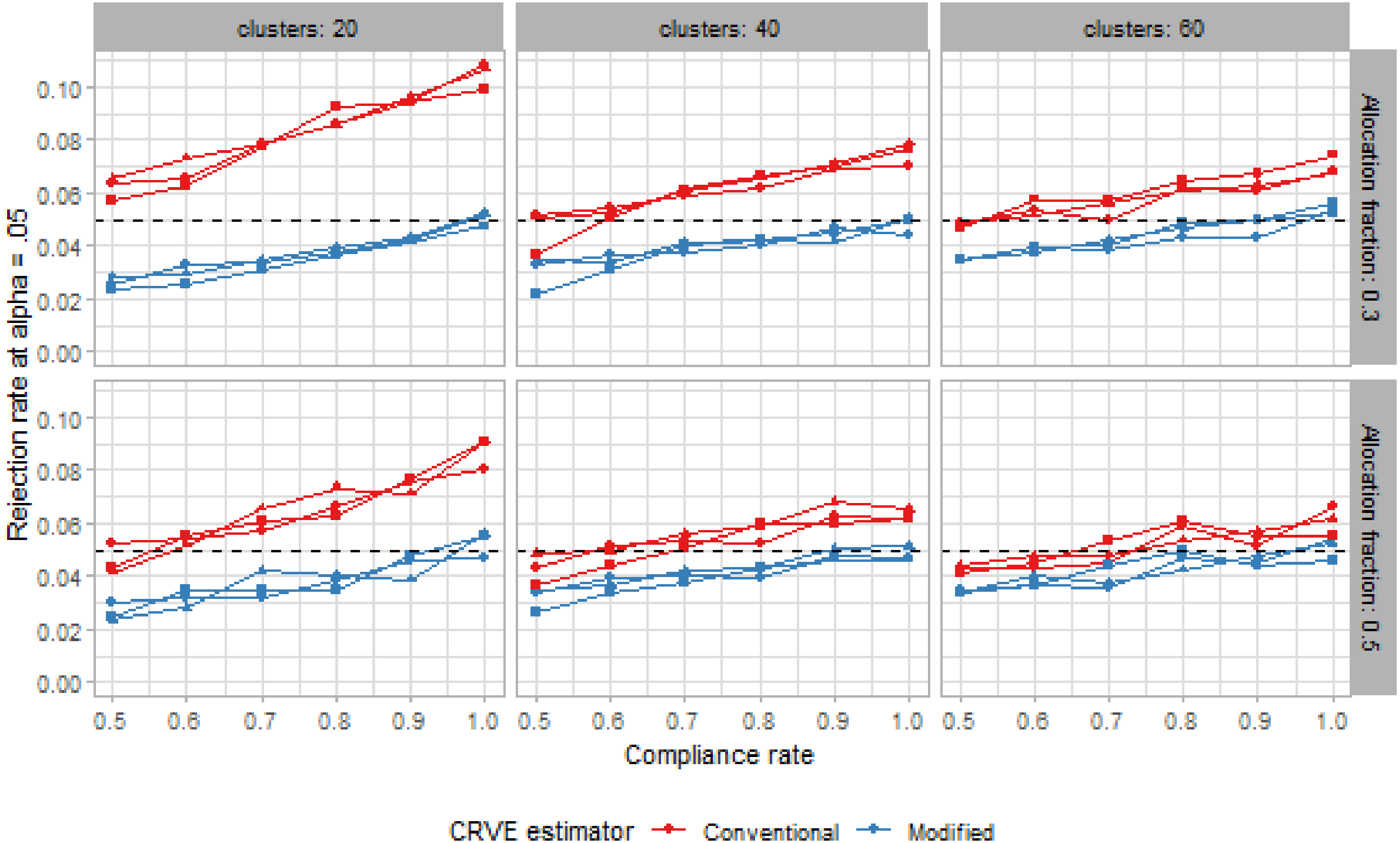
2SLS CRVE is equivalent to the delta-method estimator:

$$V(\hat{\delta}) \approx \frac{1}{\hat{\gamma}^2}\left[V(\hat{\beta}) + \hat{\delta}^2 V(\hat{\gamma}) - 2\hat{\delta} V(\hat{\beta}, \hat{\gamma})\right]$$

Using the proposed adjustment matrices gives **exactly unbiased estimates *of each component*** in the delta-method approximation, under certain working models for $(\mathbf{u}_j, \mathbf{v}_j)$.

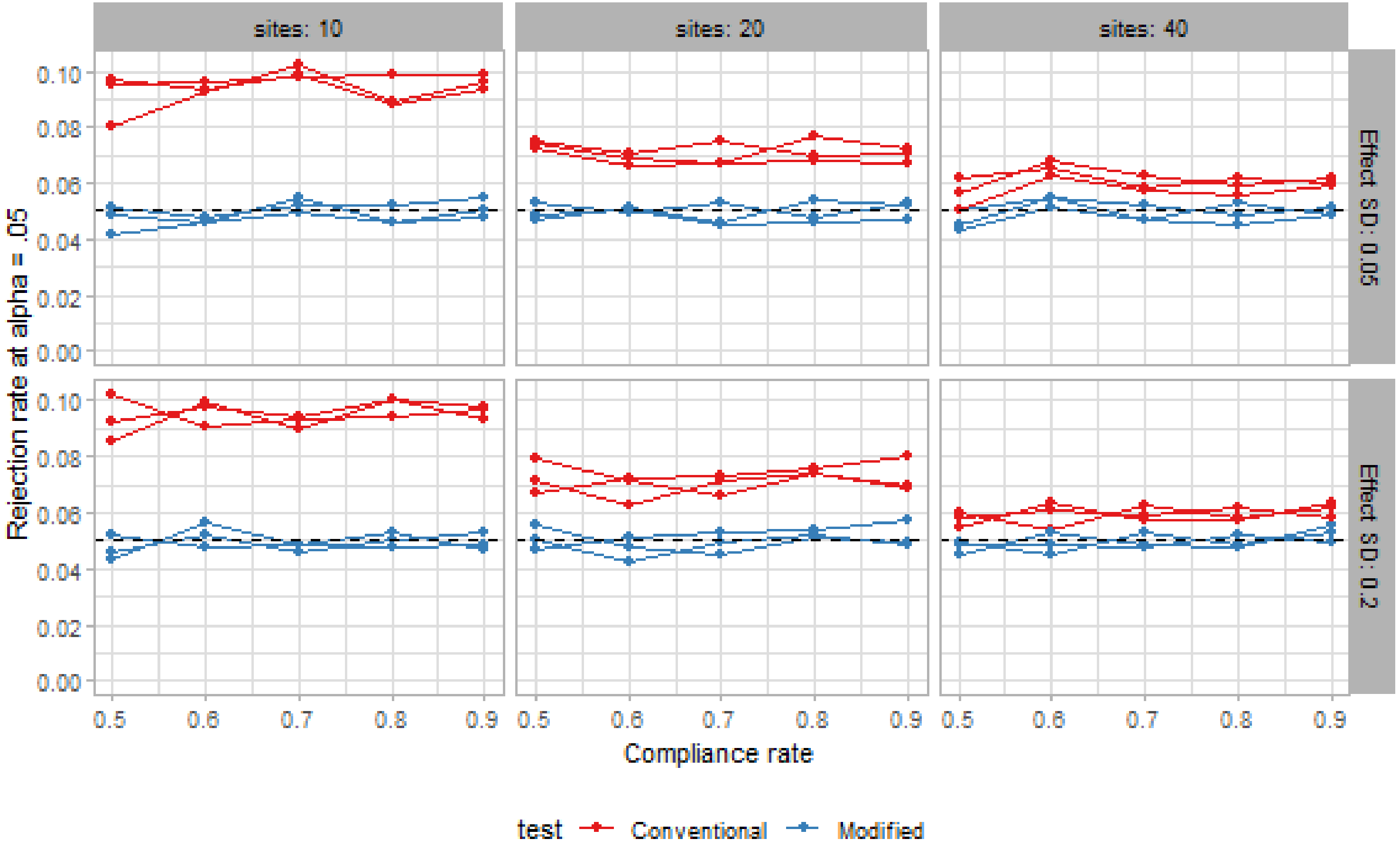
Simulations: Cluster-randomized trial

Cluster-level non-compliance



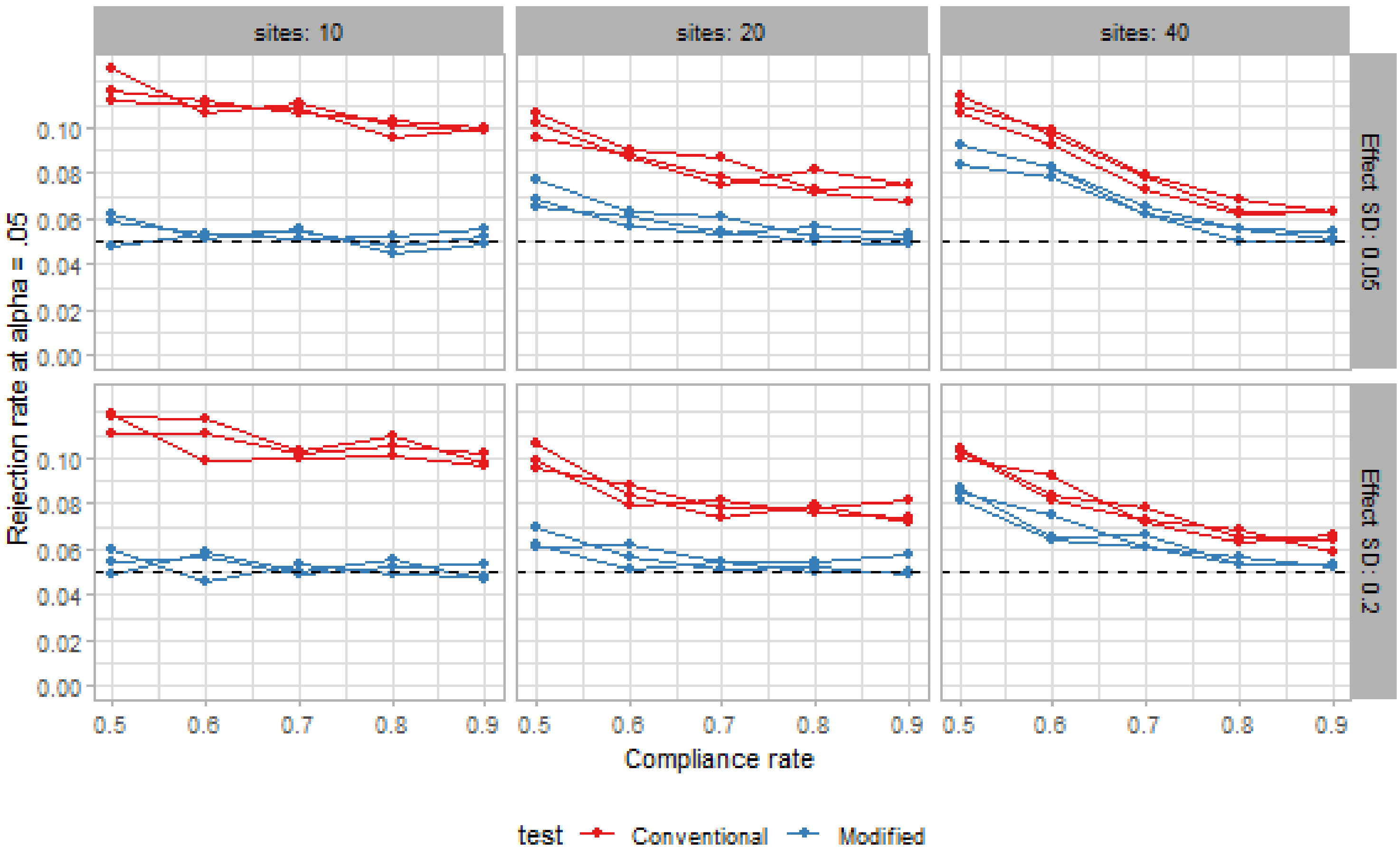
Simulations: Multi-site trial

Individual-level non-compliance, single instrument



Simulations: Multi-site trial

Individual-level non-compliance, site-specific instruments



Conclusions

- Methods implemented in **`clubSandwich`** package for R.
  - Works with `AER::ivreg`.
- Use small-sample adjusted CRVE for estimating CATE
  - In cluster-randomized trials
  - In multi-site trials with strong, single-instrument
- Future work needed on methods for weak instrument/many-instrument settings.

Contact

James E. Pustejovsky

The University of Texas at Austin

pusto@austin.utexas.edu

https://jepusto.com