AN EVALUATION OF PARAMETRIC AND NONPARAMETRIC VARIANCE ESTIMATORS IN COMPLETELY RANDOMIZED EXPERIMENTS

by

Stanley A. Lubanski

and

Peter M. Steiner

UNIVERSITY OF WISCONSIN-MADISON

2018

Background

To make valid inferences about population parameters, researchers must be able to accurately estimate those parameters, but also be able to estimate the variability of those estimates so that they may rule out sampling uncertainty as a reasonable explanation for the results observed in their sample. However, even in the simple, well-understood context of the completely randomized experiment, advice for selecting a variance estimator can vary: some recommend nonparametric, design-based estimators (Imbens & Rubin, 2015; Schochet, 2010), while others recommend catch-all corrections to the conventional parametric variance estimators (Angrist & Pischke, 2008; Long & Ervin, 2000; MacKinnon & White, 1985). In this paper, an answer is sought to the following question: do nonparametric variance estimators perform better than conventional or corrected variance estimators in completely randomized experiments (CREs) and, if so, under what conditions?

Objective

This paper focuses on evaluating the performance of four variance estimators in a CRE with two experimental groups. The first estimator will be the conventional regression variance estimator which assumes homoscedasticity of variance (i.e., constant error variance between the groups; hence this estimator is labeled "Constant"). Two design-based estimators—labeled "Neyman" and "Fixed Sample Upper Bound"; "FSUB" for short (Imbens & Rubin, 2015; Schochet, 2015)—and one corrected parametric estimator—the "HC3" estimator (MacKinnon & White, 1985)—all of which relax the homoscedasticity assumption, will be compared to the conventional estimator. Of interest is how each variance estimator will handle different sources and types of variability. Formulae are presented in Table 1 below.

By design, point estimates from a CRE will have random variability two sources: random assignment or random sampling. If a researcher wishes to generalize back to a larger target population, then they often use the "super population" model where both random assignment and random sampling are present. If a researcher is only concerned with the single sample they have drawn, then they often use what we will call the "fixed sample" model where only random assignment is present. Additional factors that impact the variability of the point estimators are also considered, including: heteroscedasticity, treatment group proportion, and covariance adjustment.

Table 1

Variance Estimators to be evaluated.

No.	Model type	Name	Notation	Point Estimator	Formula
1	Unadjusted	Constant	$\widehat{V}_{Con}$	$\hat{\tau}$	$\left(s_t^2 \frac{(m-1)}{n-2} + s_c^2 \frac{(n-m-1)}{n-2} \right) \frac{n}{m(n-m)}$
2	Unadjusted	Heteroscedasticity-Consistent Covariance Estimator 3 (HC3)	$\widehat{V}_{HC3}$	$\hat{\tau}$	$\frac{s_t^2}{m-1} + \frac{s_c^2}{n-m-1}$
3	Unadjusted	Neyman	$\widehat{V}_{Ney}$	$\hat{\tau}$	$\frac{s_t^2}{m} + \frac{s_c^2}{n-m}$
4	Unadjusted	Fixed Sample Upper Bound (FSUB)	$\widehat{V}_{FSUB}$	$\hat{\tau}$	$\frac{s_t^2}{m} + \frac{s_c^2}{n-m} - \frac{(s_t - s_c)^2}{n}$
5	Covariance Adjusted (X)	Constant w/covariates*	$\hat{V}_{Con}^X$	$\hat{\tau}^X$	$MSE(\boldsymbol{M}'\boldsymbol{M})^{-1}$
8	Covariance Adjusted (X)	HC3 w/covariates*	$\hat{V}_{HC3}^X$	$\hat{\tau}^X$	$(\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}' diag\left[\frac{(Y_i - \hat{Y}_i)^2}{(1-h_{ii})^2}\right]\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{M})^{-1}$
6	Covariance Adjusted (X)	Neyman w/covariates	$\hat{V}_{Ney}^X$	$\hat{\tau}^X$	$\frac{MSE_t}{m} + \frac{MSE_c}{n-m}$
7	Covariance Adjusted (X)	FSUB w/covariates	$\hat{V}_{FSUB}^X$	$\hat{\tau}^X$	$\frac{MSE_t}{m} + \frac{MSE_c}{n-m} - \frac{(\sqrt{MSE_t} - \sqrt{MSE_c})^2}{n}$

**formulas given for these estimators are in matrix notation, so the value of interest would be the (2, 2) element of the resulting matrix*
Note 1. **M** *is the design matrix with n rows and k+1 columns, where k is the number of predictors*
Note 2. MSE is the estimated mean square error of the residuals
Note 3. MSE_t *and* MSE_c *are the sum of squares of the residuals for the treatment and control group, respectively, divided by the appropriate degrees of freedom*
Note 4. For simplicity, "Covariance Adjusted" can refer to either the main effects or interaction model; variance estimator equations for both models are identical, but differ only in the value of the residuals, hat values, or design matrix

Research Design

In R, we ran a simulation with 10,000 iterations for each of the 24 combinations of the three following factors: inference population (four levels), heteroscedasticity (three levels), and treatment group proportion (2 levels). The base simulation used normally-distributed error terms and sample size of 1,000, but the full results will include non-normally distributed error terms and sample size of 100. For each iteration, point and variance estimates are computed for three regression models: (1) unadjusted model – the only predictor is the binary treatment indicator, Z, (2) main effects model – in a addition to Z, four covariates (Xs) are included as predictors, (3) interaction model – in addition to the main effects of Z and the Xs, the interaction of each of the Xs with Z are included as predictors. In all regression models, the predictors were centered at their sample means. Detail of the full data generating model are presented in Table 2 below.

Table 2

Data Generating Model

Variables	Explanation	Value/Model
τ	Population Average Treatment Effect (PATE)	5
$\mathbf{X}$	Covariates matrix	$\sim N_4\left(\boldsymbol{\mu}_x = \begin{pmatrix}1\\2\\1\\2\end{pmatrix},\ \Sigma_x = \begin{pmatrix}1 & .4 & .2 & 0\\ .4 & 1 & 0 & .1\\ .2 & 0 & 1 & 0\\ 0 & .1 & 0 & 1\end{pmatrix}\right)$
$\boldsymbol{\beta}_{\mathbf{y}}$	Vector of coefficients relating X's to Y(0)	$\begin{pmatrix}\beta_1 = 4\\ \beta_2 = 2\\ \beta_3 = 1\\ \beta_4 = 3\end{pmatrix}$
ϵ_y	Y(0) error term	$\sim N(0, 25)$
$Y(0)$	Control Potential Outcome	$\boldsymbol{X\beta_y} + \epsilon_y$
ϵ_{ran}	$Y(1)_{ran}$ error term	$\sim N(0, \sigma^2_{Y(0)})$
$Y(1)_{con}$	Treatment Potential Outcome (constant)	$Y(0) + \tau$
$Y(1)_{ran}$	Treatment Potential Outcome (random)	$Y(0) + \tau + \epsilon_{ran}$
$Y(1)_{sys}$	Treatment Potential Outcome (systematic)	$Y(0) + \tau X_1$

Data Analysis

To determine if the variance estimators were biased, the mean variance estimates were compared to the empirical variance of the point estimate across simulations, in a ratio of the estimated variance over the "true" variance. However, to account for the finiteness of the simulation, the variance ratios were bootstrapped 1,000 times and then 99% bootstrap confidence intervals were computed. If the confidence intervals contained 1, then the estimator was considered unbiased, else, it was considered biased.

Results

The plot of the results for the fixed sample inference population are presented in Figure 1; results for other inference populations will be discussed in relation to this plot. Taking the figure row-by-row, first is the case of no treatment effect heterogeneity. In the left plot the proportion of treated units, "p", equals 0.5 and all estimators are unbiased. In the right plot, where $p = 0.1$, the only bias occurs in the interaction model where small biases exist for the nonparametric and HC3 estimators. Super population results for this row are essentially identical.

In the second row with random treatment effect heterogeneity, in the left plot, all estimators are upwardly biased because, for fixed samples, the presence of treatment effect heterogeneity reduces point estimator variability (Imbens & Rubin, 2015; Schochet, 2010). However, only the FSUB estimator has some correction to reflect this and thus its bias is less severe. In the right plot where $p = 0.1$, the Constant estimator becomes severely biased in all three regression models, underestimating the variance by more than 50% in the covariance adjusted models. Some of the other estimators are slightly upwardly biased. The upward bias is not present in the super population model, so there the FSUB estimator is downwardly biased; the rest of the results are nearly identical.

Finally, in the third row with systematic treatment effect heterogeneity, when $p = 0.5$, the estimators differ little, but some upward bias is present, as in the previous row of plots. None of the estimators are biased when correctly modeling the heterogeneity. When $p = 0.1$, again, the Constant estimator is downwardly biased (except when the model is correct), but the other estimators are similar. Again, the upward bias is not present in the super population, but due to sample mean centering, there is downward bias for the covariance adjusted model with interactions.

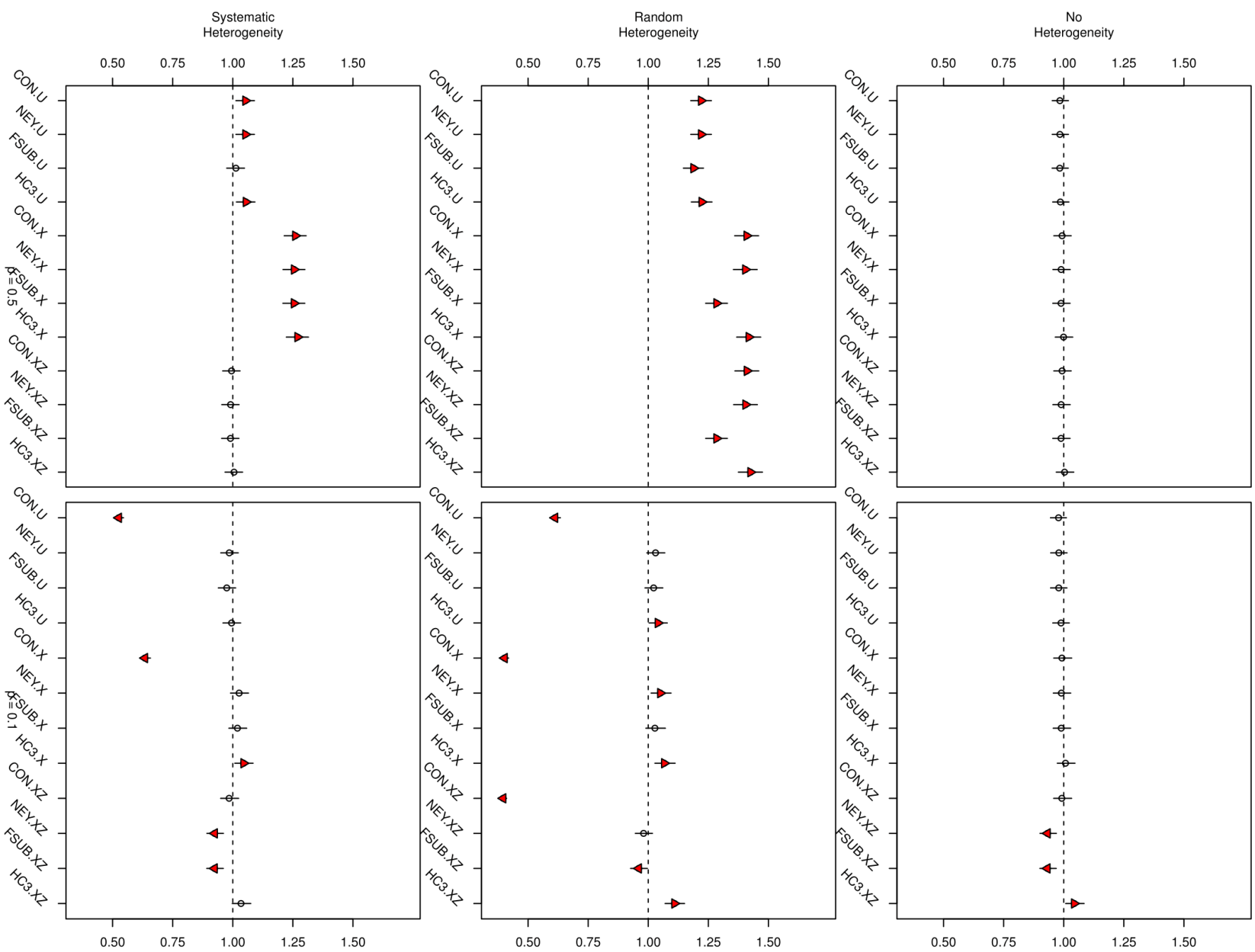
Systematic Heterogeneity
Random Heterogeneity
No Heterogeneity
0.50
0.75
1.00
1.25
1.50
CON.U
NEY.U
FSUB.U
HC3.U
CON.X
NEY.X
FSUB.X
HC3.X
CON.XZ
NEY.XZ
FSUB.XZ
HC3.XZ
β=0.5
β=0.1

Figure 1. *Variance estimator bias as variance ratios (estimated variance over true variance) for four variance estimators*. Rows indicate level of treatment effect heterogeneity (top-to-bottom: none, random, systematic). Columns indicate treated unit proportion (left-to-right: $p = 0.5$, $p = 0.1$). X-axis of plots refer to variance estimators; first four estimates from unadjusted regression model; next four from covariance adjusted regression model (main effects only); last four from covariance adjusted model (main effects plus interactions). Estimators with red triangles indicate bias, i.e., that the 99% bootstrap interval does not include 1.

Conclusions

If the CRE is balanced and there is little or no heteroscedasticity, then all variance estimators work well, but if the design is unbalanced, then the conventional parametric estimator is best as it pools the variance from both treatment groups to give the most efficient variance estimate. However, if there is random, idiosyncratic variability causing heteroscedasticity, then the nonparametric estimators are best, especially if the design is unbalanced or if working under the fixed sample model. Finally, if the heteroscedasticity is caused by systematic variation, then the choice of variance estimator matters little and it is most important to model the variation appropriately and to avoid sample-mean-centering if working under the super population model.

References

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Long, J. S., & Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *American Statistician*, *54*(3), 217–224. https://doi.org/10.2307/2685594

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325.

Schochet, P. Z. (2010). Is regression adjustment supported by the Neyman model for causal inference? *Journal of Statistical Planning and Inference*, *140*(1), 246–259. https://doi.org/10.1016/j.jspi.2009.07.008

Schochet, P. Z. (2015). Statistical Theory for the RCT-YES Software: Design-Based Causal Inference for RCTs, (June). Retrieved from https://files.eric.ed.gov/fulltext/ED556496.pdf