Erratum: Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models

James E. Pustejovsky *
University of Wisconsin - Madison
and
Elizabeth Tipton †
Northwestern University

September 29, 2022

*Department of Educational Psychology, University of Wisconsin - Madison, 1025 West Johnson Street, Madison, WI 53706. Email: pustejovsky@wisc.edu

†Department of Statistics, Northwestern University. Email: tipton@northwestern.edu

Pustejovsky and Tipton (2018) considered how to do cluster-robust variance estimation in fixed effects models estimated by weighted (or unweighted) least squares. Theorem 2 of the paper concerns a computational short cut for a certain cluster-robust variance estimator in models with cluster-specific fixed effects. The theorem is incorrect as stated. In this correction, we review the CR2 variance estimator, describe the assertion of the theorem as originally stated, and explain the error. We then provide a revised version of the theorem.

1 A fixed effects model

For data that can be grouped into m clusters of observations, Pustejovsky and Tipton (2018) considered the model

$$\mathbf{y}_i = \mathbf{R}_i\boldsymbol{\beta} + \mathbf{S}_i\boldsymbol{\gamma} + \mathbf{T}_i\boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \tag{1}$$

where $\mathbf{y}_i$ is an $n_i \times 1$ vector of responses for cluster i, $\mathbf{R}_i$ is an $n_i \times r$ matrix of focal predictors, $\mathbf{S}_i$ is an $n_i \times s$ matrix of additional covariates that vary across multiple clusters, and $\mathbf{T}_i$ is an $n_i \times t$ matrix encoding cluster-specific fixed effects, all for $i = 1, ..., m$. The cluster-specific fixed effects satisfy $\mathbf{T}_h\mathbf{T}_i' = \mathbf{0}$ for $h \neq i$. Interest centers on inference for the coefficients on the focal predictors $\boldsymbol{\beta}$.

Pustejovsky and Tipton (2018) considered estimation of Model @ref(eq:regression) by weighted least squares (WLS). Let $\mathbf{W}_1, ..., \mathbf{W}_m$ be a set of symmetric weight matrices used for WLS estimation. The CR2 variance estimator involves specifying a working model for the structure of the errors. Consider a working model $\text{Var}\left(\boldsymbol{\epsilon}_i|\mathbf{R}_i, \mathbf{S}_i, \mathbf{T}_i\right) = \sigma^2\boldsymbol{\Phi}_i$ where $\boldsymbol{\Phi}_i$ is a symmetric $n_i \times n_i$ matrix that may be a function of a low-dimensional, estimable parameter. Under some circumstances, the weight matrices might be taken as $\mathbf{W}_i = \hat{\boldsymbol{\Phi}}_i^{-1}$, where $\hat{\boldsymbol{\Phi}}_i$ is an estimate of $\boldsymbol{\Phi}_i$.

2 The CR2 variance estimator

Pustejovsky and Tipton (2018) provided a generalization of the bias-reduced linearization estimator introduced by McCaffrey, Bell, and Botts (2001) and Bell and McCaffrey (2002) that can be applied to Model @ref(eq:regression), referred to as CR2. In order to define the CR2 variance estimator, some further notation will be required, just as defined in Pustejovsky and Tipton (2018). Let $N = \sum_{i=1}^m n_i$ be the total sample size. Let $\mathbf{U}_i = [\mathbf{R}_i \; \mathbf{S}_i]$ be the set of predictors that vary across clusters and $\mathbf{X}_i = [\mathbf{R}_i \; \mathbf{S}_i \; \mathbf{T}_i]$ be the full set of predictors. Let $\mathbf{R}$, $\mathbf{S}$, $\mathbf{T}$, $\mathbf{U}$, and $\mathbf{X}$ denote the stacked versions of the cluster-specific matrices (i.e., $\mathbf{R} = [\mathbf{R}_1' \; \mathbf{R}_2' \; \cdots \; \mathbf{R}_m']'$, etc.). Let $\mathbf{W} = \bigoplus_{i=1}^m \mathbf{W}_i$ and $\mathbf{\Phi} = \bigoplus_{i=1}^m \mathbf{\Phi}_i$. For a generic matrix $\mathbf{Z}$, let $\mathbf{M}_Z = (\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1}$ and $\mathbf{H}_\mathbf{Z} = \mathbf{Z}\mathbf{M}_\mathbf{Z}\mathbf{Z}'\mathbf{W}$. Let $\mathbf{C}_i$ be the $n_i \times N$ matrix that selects the rows of cluster i from the full set of observations, such that $\mathbf{X}_i = \mathbf{C}_i\mathbf{X}$. Finally, let $\mathbf{D}_i$ be the upper-right Cholesky factorization of $\mathbf{\Phi}_i$.

These operators provide a means to define absorbed versions of the predictors. Let $\ddot{\mathbf{S}} = (\mathbf{I} - \mathbf{H}_\mathbf{T})\,\mathbf{S}$ be the covariates after absorbing the cluster-specific effects, let $\ddot{\mathbf{U}} = (\mathbf{I} - \mathbf{H}_\mathbf{T})\,\mathbf{U}$ be an absorbed version of the focal predictors and the covariates, and let $\ddot{\mathbf{R}} = (\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{S}}})\,(\mathbf{I} - \mathbf{H}_\mathbf{T})\,\mathbf{R}$ be the focal predictors after absorbing the covariates and the cluster-specific fixed effects.

With this notation established, the CR2 variance estimator has the form

$$\mathbf{V}^{CR2} = \mathbf{M}_{\ddot{\mathbf{R}}} \left(\sum_{i=1}^m \ddot{\mathbf{R}}_i' \mathbf{W}_i \mathbf{A}_i \mathbf{e}_i \mathbf{e}_i' \mathbf{A}_i \mathbf{W}_i \ddot{\mathbf{R}}_i \right) \mathbf{M}_{\ddot{\mathbf{R}}}, \tag{2}$$

where $\ddot{\mathbf{R}}_i = \mathbf{C}_i \ddot{\mathbf{R}}$ is the cluster-specific matrix of absorbed focal predictors, $\mathbf{e}_i$ is the vector of weighted least squares residuals from cluster i, and $\mathbf{A}_1, ..., \mathbf{A}_m$ are a set of adjustment matrices that correct the bias of the residual cross-products. The adjustment matrices are calculated as follows. Define the matrices

$$\mathbf{B}_i = \mathbf{D}_i \mathbf{C}_i \left(\mathbf{I} - \mathbf{H}_{\mathbf{X}}\right) \boldsymbol{\Phi} \left(\mathbf{I} - \mathbf{H}_{\mathbf{X}}\right)' \mathbf{C}_i' \mathbf{D}_i' \tag{3}$$

for $i = 1, ..., m$. The adjustment matrices are then calculated as

$$\mathbf{A}_i = \mathbf{D}_i' \mathbf{B}_i^{+1/2} \mathbf{D}_i, \tag{4}$$

where $\mathbf{B}_i^{+1/2}$ is the symmetric square root of the Moore-Penrose inverse of $\mathbf{B}_i$. Theorem 1 of Pustejovsky and Tipton (2018) shows that, if the working model $\boldsymbol{\Phi}$ is correctly specified and some conditions on the rank of $\mathbf{U}$ are satisfied, then the CR2 estimator is exactly unbiased for the sampling variance of the weighted least squares estimator of $\boldsymbol{\beta}$. Although it is defined based on a working model, the CR2 estimator remains close to unbiased and outperforms alternative sandwich estimators even when the working model is not correctly specified.

3 The original statement of Theorem 2

The adjustment matrices given in @ref(eq:A-matrix) can be expensive to compute directly because the $\mathbf{B}_i$ matrices involve computing a "residualized" version of the $N \times N$ matrix $\boldsymbol{\Phi}$ involving the full set of predictors $\mathbf{X}$—including the cluster-specific fixed effects $\mathbf{T}_1, ..., \mathbf{T}_m$. Theorem 2 considered whether one can take a computational short cut by omitting the cluster-specific fixed effects from the calculation of the $\mathbf{B}_i$ matrices. Specifically, define the modified matrices

$$\tilde{\mathbf{B}}_i = \mathbf{D}_i \mathbf{C}_i \left(\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}}\right) \boldsymbol{\Phi} \left(\mathbf{I} - \mathbf{H}_{\ddot{\mathbf{U}}}\right)' \mathbf{C}_i' \mathbf{D}_i' \tag{5}$$

and

$$\tilde{\mathbf{A}}_i = \mathbf{D}_i' \tilde{\mathbf{B}}_i^{+1/2} \mathbf{D}_i. \tag{6}$$

Theorem 2 claimed that if the weight matrices are inverse of the working model, such that $\mathbf{W}_i = \mathbf{\Phi}_i^{-1}$ for $i = 1, ..., m$, then $\tilde{\mathbf{B}}_i^{+1/2} = \mathbf{B}_i^{+1/2}$ and hence $\tilde{\mathbf{A}}_i = \mathbf{A}_i$. The implication is that the cluster-specific fixed effects can be ignored when calculating the adjustment matrices. However, the claimed equivalence does not actually hold. The proof of Theorem 2 as given in Pustejovsky and Tipton (2018) relied on a Woodbury identity for generalized inverses that does not hold for $\mathbf{B}_i$ because the rank conditions are not satisfied.

4 A revised Theorem 2

References

Bell, Robert M, and Daniel F McCaffrey. 2002. "Bias reduction in standard errors for linear regression with multi-stage samples." *Survey Methodology* 28 (2): 169–81.

McCaffrey, Daniel F, Robert M Bell, and Carsten H Botts. 2001. "Generalizations of biased reduced linearization." In *Proceedings of the Annual Meeting of the American Statistical Association*. 1994.

Pustejovsky, James E., and Elizabeth Tipton. 2018. "Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models." *Journal of Business & Economic Statistics* 36 (4): 672–83. `https://doi.org/10.1080/07350015.2016.1247004`.