Robust Standard Errors in Small Samples

Michal Kolesár

August 16, 2019

Contents

1 Description

```r
library(dfadjust)
```

The implementation, described below, allows for fixed effects and for large clusters. Example:

2 Methods

This section describes the implementation of the Imbens and Kolesár [2016] and Bell and McCaffrey [2002] degrees of freedom adjustments.

There are S clusters, and we observe n_s observations in cluster s, for a total of $n = \sum_{s=1}^{S} n_s$ observations. We handle the case with independent observations by letting each observation be in its own cluster, with $S = n$. Consider the linear regression of a scalar outcome Y_i onto a p-vector of regressors X_i,

$$Y_i = X_i'\beta + u_i, \qquad E[u_i \mid X_i] = 0.$$

We're interested in inference on $\ell'\beta$ for some fixed vector $\ell \in \mathbb{R}^p$. Let X, u, and Y denote the design matrix, and error and outcome vectors, respectively. For any $n \times k$ matrix M, let M_s denote the $n_s \times k$ block corresponding to cluster s, so that, for instance, Y_s corresponds to the outcome vector in cluster s. For a positive semi-definite matrix M, let $M^{1/2}$ be a matrix satisfying $M^{1/2'}M^{1/2} = M$, such as its symmetric square root or its Cholesky decomposition.

Assume that

$$E[u_s u_s' \mid X] = \Omega_s, \quad \text{and} \quad E[u_s u_t' \mid X] = 0 \quad \text{if } s \neq t.$$

Denote the conditional variance matrix of u by Ω, so that Ω_s is the block of Ω corresponding to cluster s. We estimate $\ell'\beta$ using OLS. In R, the OLS estimator is computed via a QR decomposition, $X = QR$, where $Q'Q = I$ and R is upper-triangular, so we can write the estimator as

$$\ell'\hat{\beta} = \ell' \left(\sum_s X_s' X_s \right)^{-1} \sum_s X_s Y_s = \tilde{\ell}' \sum_s Q_s' Y_s, \qquad \tilde{\ell} = R^{-1'} \ell.$$

It has variance

$$V := \operatorname{var}(\ell'\hat{\beta} \mid X) = \ell' \left(X'X\right)^{-1} \sum_s X_s' \Omega_s X_s \left(X'X\right)^{-1} \ell = \tilde{\ell}' \sum_s Q_s' \Omega_s Q_s \tilde{\ell}.$$

2.1 *Variance estimate*

We estimate V using a variance estimator that generalizes the HC2 variance estimator to clustering. Relative to the LZ2 estimator described in Imbens and Kolesár [2016], we use a slight modification that allows for fixed effects:

$$\hat{V} = \ell'(X'X)^{-1} \sum_s X_s' A_s \hat{u}_s \hat{u}_s' A_s' X_s (X'X)^{-1} \ell = \ell' R^{-1} \sum_s Q_s' A_s \hat{u}_s \hat{u}_s' A_s' Q_s R'^{-1} \ell = \sum_{s=1}^{S} (\hat{u}_s' a_s)^2,$$

where

$$\hat{u}_s := Y_s - X_s \hat{\beta} = u_s - Q_s Q' u, \qquad a_s = A_s' Q_s \tilde{\ell},$$

and the matrix A_s is given by the symmetric square root of the inverse of $I - Q_s Q_s'$, or else its pseudo-inverse if it is singular, as is the case, for example, if X contains fixed effects. We do not need to insist on $I - Q_s Q_s'$ to be invertible, since, using the identity

$$\hat{V} = u \sum_s (I - QQ')_s' a_s a_s' (I - QQ')_s u,$$

one can verify by simple algebra that a sufficient condition for $\hat{V}$ to be unbiased under homoskedasticity is that $Q_s' A_s (I - Q_s Q_s') A_s Q_s = Q_s' Q_s$ (see, for example, Pustejovsky and Tipton [2018], for details).

If the observations are independent, the vector of leverages $(Q_1' Q_1, \ldots, Q_n' Q_n)$ can be computed dirrectly using the `stats::hatvalues` function. In this case, use this function to compute $A_i = 1/\sqrt{1 - Q_i' Q_i}$ directly, and we then compute $a_i = A_i Q_i' \tilde{\ell}$ using vector operations. For the case with clustering, computing the spectral decomposition of $I - Q_s Q_s'$ can be expensive or even infeasible if the cluster size n_s is large. We therefore use the following result, suggested to us by Ulrich Müller, allows us to compute a_s by computing a spectral decomposition of a $p \times p$ matrix.

- Let $Q_s' Q_s = \sum_{i=1}^{p} \lambda_{is} r_{is} r_{is}'$ be the spectral decomposition of $Q_s' Q_s$. Then $A_s = \sum_{i\colon \lambda_i \neq 1} (1 - \lambda_i)^{-1/2} Q_s r_{is} r_{is}' Q_s'$, satisfies $A_s (I - Q_s Q_s') A_s = I$.

  This follows from the fact that $I - Q_s Q_s'$ has eigenvalues $1 - \lambda_{is}$ and eigenvectors $Q_s r_{is}$, and hence its pseudoinverse is $\sum_{i\colon \lambda_i\neq 1}(1-\lambda_i)^{-1}Q_s r_{is}r_{is}'Q_s'$.

Using the lemma, we can compute a_s efficiently as:

$$a_s = \sum_{i\colon \lambda_i\neq 1}(1-\lambda_i)^{-1/2}Q_s r_{is}r_{is}'Q_s'Q_s\tilde{\ell} = Q_s D_s\tilde{\ell}, \qquad D_s = \sum_{i\colon \lambda_i\neq 1}\lambda_i(1-\lambda_i)^{-1/2}r_{is}r_{is}'.$$

2.2 *Degrees of freedom correction*

Let G be an $n\times S$ matrix with columns $(I-QQ')_s'a_s$. Then the Bell and McCaffrey [2002] adjustment sets the degrees of freedom to

$$f_{\text{BM}} = \frac{\operatorname{tr}(G'G)^2}{\operatorname{tr}((G'G)^2)}.$$

Since $(G'G)_{st} = a_s'(I-QQ')_s(I-QQ)_t'a_t = a_s(\mathbb{1}\{s=t\}-Q_sQ_t')a_t$, the matrix $G'G$ can be efficiently computed as

$$G'G = \operatorname{diag}(a_s'a_s) - BB' \qquad B_{sk} = a_s'Q_{sk}.$$

Note that B is an $S\times p$ matrix, so that computing the degrees of freedom adjustment only involves $p\times p$ matrices:

$$f_{\text{BM}} = \frac{(\sum_s a_s'a_s - \sum_{s,k}B_{sk}^2)^2}{\sum_s(a_s'a_s)^2 - 2\sum_{s,k}(a_s'a_s)B_{sk}^2 + \sum_{s,t}(B_s'B_t)^2}.$$

If the observations are independent, we compute B directly as `B <- a*Q`, and since a_i is a scalar, we have

$$f_{\text{BM}} = \frac{(\sum_i a_i^2 - \sum_{sk}B_{sk}^2)^2}{\sum_i a_i^4 - 2\sum_i a_i^2 B_i'B_i + \sum_{i,j}(B_i'B_j)^2}.$$

The Imbens and Kolesár [2016] degrees of freedom adjustment instead sets

$$f_{IK} = \frac{\operatorname{tr}(G'\hat{\Omega}G)^2}{\operatorname{tr}((G'\hat{\Omega}G)^2)},$$

where $\hat{\Omega}$ is an estimate of the Moulton [1986] model of the covariance matrix, under which $\Omega_s = \sigma_\epsilon^2 I_{n_s} + \rho\iota_{n_s}\iota_{n_s}'$. Using simple algebra, one can show that in this case,

$$G'\Omega G = \sigma_\epsilon^2\operatorname{diag}(a_s'a_s) - \sigma_\epsilon^2 BB' + \rho(D-BF')(D-BF')',$$

where

$$F_{sk} = \iota_{n_s}'Q_{sk}, \qquad D = \operatorname{diag}(a_s'\iota_{n_s})$$

which can again be computed even if the clusters are large. The estimate $\hat{\Omega}$ replaces σ_ϵ^2 and ρ with analong estimates.

References

Robert M. Bell and Daniel F. McCaffrey. Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28(2):169–181, 2002.

Guido W. Imbens and Michal Kolesár. Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98(4):701–712, October 2016. doi: 10.1162/REST_a_00552.

Brent R. Moulton. Random group effects and the precision of regression estimates. *Journal of Econometrics*, 32(3):385–397, 1986.

James E. Pustejovsky and Elizabeth Tipton. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4):672–683, October 2018. doi: 10.1080/07350015.2016.1247004.