Standard errors with very small or very large datasets

Michal Kolesár

ECO539B, Spring 2025

March 23, 2025

Problems with standard inference

Degrees of freedom correction

Alternative alternatives

- Consider regression of Y_i onto $X_i = (D_i, W_i)$, $k = \dim(X_i)$. When are Eicker-Huber-White (EHW) and Liang-Zeger (LZ) standard errors unreliable?
- Suppose $E[Y_i \mid X_i = x] = d\beta + w'\gamma$, and that we want to do inference on β conditional on X
  - Perhaps not the best approach for ensuring causal or descriptive interpretation for β robust to non-linearity of regression function.
  - But it makes it easy to think through hiccups with standard inference.

Regularity conditions for central limit theorem (CLT)

No fat tails $E[\epsilon_i^{2+\eta} \mid X]$ is bounded for some $\eta > 0$.

Low partial leverage $\max_i H_{\ddot{D},ii} \to 0$ (i.e. no outliers in X_i).

Regularity conditions for inference

For consistency of $\hat{V}_{\text{EHW}}$ need to strengthen the leverage condition. Sufficient conditions:

Leverage for inference Either $k \max_i H_{X,ii} \to 0$ or $\sqrt{nk} \max_i H_{\ddot{D},ii} \to 0$.

- First version ensures consistency of full regression function. Violated with fixed effects. Conditions allow for "high-dimensional" setting with $k \to \infty$, but trouble if $k \asymp n$.
- Research question: how necessary are these conditions?

Lemma

Suppose that the above conditions hold. Then EHW standard errors lead to asymptotically valid inference.

What can go wrong?

1. CLT fails
    - Look at outliers. Can winsorize, but changes interpretation of β.
        - Alternative inference procedures exist (Müller 2023).
    - Look at partial leverage $\max_i H_{\ddot{D},ii}$
2. EHW variance estimator is not consistent: it displays finite-sample bias or substantial sampling variability: t-stats not normal, leading to undercoverage.
    - Natural solution is to bias-correct estimator and use degrees of freedom (DoF) correction

Example 1

Suppose that $D_1 = C\sqrt{n}$ for some constant C, while $D_i = 1$ if $i > 1$, and that $W_i = 1$. Then $H_{X,11} = 1$, while $H_{X,ii} = 1/(n-1)$ for $i > 1$. $\hat{\beta}$ is $\sqrt{n}$-consistent, but not asymptotically normal unless ϵ_1 happens to be normal.

Bo Honoré's outlier detection method

Wish to check whether first observation outlier, so set $D_i = \mathbb{1}\{i = 1\}$. W_i are well-behaved controls. Then (i) $H_{X,11} = 1$ (ii) $\hat{\epsilon}_1 = 0$ (iii) $\hat{\gamma}$ consistent, and $\hat{\beta}$ converges to $\beta + \epsilon_1$, and (iv) t-statistic for $\hat{\beta}$ based on EHW standard errors converges to $\pm\infty$ irrespective of value of β.

Berhens-Fisher problem (Behrens 1929; Fisher 1939)

n_1 observations are treated, n_0 controls. Only covariates are intercept. Assume $\epsilon_i \mid D_i \sim \mathcal{N}(0, \sigma^2(D_i))$. Clear that even if n large, effective sample size small if $\min\{n_1, n_0\}$ is small. Leverage reflects this: $H_{X,ii} = 1/n_{D_i}$.

- More complicated version of this problem arises in differences-in-differences contexts with a few treated observations.

Clustering

Clustering introduces additional complications:

- Sample size is determined by number of clusters S: asymptotics are as $S \to \infty$.
- Rate of convergence depends heterogeneity in cluster sizes and on within-cluster correlation structure. It'll be at most $n^{-1/2}$, but it can be even much slower than $S^{-1/2}$.
- Necessary that $\max_i H_{\ddot{D},ii} \to 0$ for CLT to hold. But what matters is leverage of whole cluster, so sufficient leverage condition substantially stronger.

Problems with standard inference

Degrees of freedom correction

Alternative alternatives

Bias of EHW

In general:

$$B = E[\hat{V}_{\text{EHW},11} \mid X] - \mathcal{V}_{\text{cx},11} = \frac{\sum_i E[\hat{\epsilon}_i^2 - \sigma^2(X_i) \mid X_i]\ddot{D}_i^2}{(\sum_i \ddot{D}_i^2)^2} \asymp \frac{\sum_i H_{X,ii}\ddot{D}_i^2}{(\sum_i \ddot{D}_i^2)^2}$$

Under homoskedastic errors:

$$E[\hat{V}_{\text{EHW}} \mid X] - \mathcal{V}_{dc} = \sigma^2 n(X'X)^{-1} \sum_i (H_{X,ii} - 1)X_i X_i'(X'X)^{-1} \leq 0.$$

Simple solution is to replace EHW with (MacKinnon and White 1985)

$$\hat{V}_{HC2} = n(X'X)^{-1} \sum_i \frac{\hat{\epsilon}_i^2}{1 - H_{X,ii}} X_i X_i'(X'X)^{-1},$$

Variance of variance estimator

- Using HC2 estimator solves bias issue, but another issue is variance: reason for using t-distribution critical values under homoskedastic normal errors.
- Makes sense to also use DoF correction with heteroskedastic errors.
- Simplest approach is to use ν DoF, where ν matches first 2 moments of variance estimator under homoskedasticity (Satterthwaite 1946)
  - Formula in lecture notes
- Key point: DoF adjustment reflects distribution of covariates and hence any leverage issues

- Similar adjustments can be applied under clustering.
- Bell and McCaffrey (2002) generalize both bias and DoF correction, based on matching DoF under homoskedastic Gaussian benchmark
- But can use other working models. See Imbens and Kolesár (2016) for details, and Hansen (2021) for refinement.
- These adjustments are heuristic, but tend to work well in practice.
    - Be on a lookout for a working paper that formalizes these heuristics.

Problems with standard inference

Degrees of freedom correction

Alternative alternatives

- Possible to construct variance estimators that are exactly unbiased.
- Approach 1: use Hadamard products (Dobriban and Su 2024; Cattaneo, Jansson, and Newey 2018). Involves inverting $n \times n$ matrices.
- Leave-out approach (Kline, Saggio, and Sølvsten 2020; Jochmans 2022): estimate $\sigma^2(X_i)$ in variance formula not by $(Y_i - X_i'\hat{\theta})^2$ used by EHW, but by unbiased estimator

$$\check{\sigma}_i^2 = Y_i(Y_i - X_i'\hat{\theta}_{-i}) = \frac{Y_i(Y_i - X_i'\hat{\theta})}{1 - H_{X,ii}}$$

  Downside: can be noisy
- Formally, both approaches with when $p \asymp n$.

Wild bootstrap

- Popularized by Cameron, Gelbach, and Miller (2008).
- Confidence intervals have to be computed by test inversion:
  1. To test the null $\ell'\theta = c$, compute the OLS estimate θ subject to this restriction, obtaining the restricted estimate $\hat{\theta}_r$ and residuals $\hat{\epsilon}_i^r$.
  2. Let $Y_i^* = X_i'\hat{\theta}^r + g_{s(i)}^*\hat{\epsilon}_i^r$, where $g_{s(i)}^* \in \{-1, 1\}$ (with equal probability), and let $X_i^* = X_i$. Compute $\hat{\theta}^*$ using ordinary least squares (OLS) in this bootstrap sample.
  3. As a critical value for the test statistic $|\ell'\hat{\theta} - c|$, use the $1 - \alpha$ quantile of $|\ell'(\hat{\theta}^* - \hat{\theta}^r)|$
- Canay, Santos, and Shaikh (2021) show formally that this method works even with a fixed number of clusters, but need strong homogeneity conditions on distro on covariates across clusters

References i

Behrens, Walter Ulrich. 1929. “Ein Beitrag Zur Fehlerberechnung Bei Wenigen Beobachtungen.” *Landwirtschaftliche Jahrbücher* 68:807–837.

Bell, Robert M., and Daniel F. McCaffrey. 2002. “Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples.” *Survey Methodology* 28, no. 2 (December): 169–181. https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X20020029058.

Cameron, Colin A., Jonah B. Gelbach, and Douglas L. Miller. 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *The Review of Economics and Statistics* 90, no. 3 (August): 414–427. https://doi.org/10.1162/rest.90.3.414.

Canay, Ivan Alexis, Andres Santos, and Azeem M Shaikh. 2021. “The Wild Bootstrap with a “Small” Number of “Large” Clusters.” *Review of Economics and Statistics* 103, no. 2 (May): 346–363. https://doi.org/10.1162/rest_a_00887.

Cattaneo, Matias D., Michael Jansson, and Whitney K. Newey. 2018. “Inference in Linear Regression Models with Many Covariates and Heteroscedasticity.” *Journal of the American Statistical Association* 113, no. 523 (July): 1350–1361. https://doi.org/10.1080/01621459.2017.1328360.

Dobriban, Edgar, and Weijie J. Su. 2024. “Robust Inference Under Heteroskedasticity via the Hadamard Estimator,” January. arXiv: 1807.00347.

Fisher, Ronald Aylmer. 1939. “The Comparison of Samples with Possibly Unequal Variances.” *Annals of Eugenics* 9, no. 2 (June): 174–180. https://doi.org/10.1111/j.1469-1809.1939.tb02205.x.

Hansen, Bruce E. 2021. “The Exact Distribution of the White T-Ratio.” Working paper, University of Wisconsin.

Imbens, Guido W., and Michal Kolesár. 2016. "Robust Standard Errors in Small Samples: Some Practical Advice." *Review of Economics and Statistics* 98, no. 4 (October): 701–712. https://doi.org/10.1162/REST_a_00552.

Jochmans, Koen. 2022. "Heteroscedasticity-Robust Inference in Linear Regression Models With Many Covariates." *Journal of the American Statistical Association* 117, no. 538 (April): 887–896. https://doi.org/10.1080/01621459.2020.1831924.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-Out Estimation of Variance Components." *Econometrica* 88, no. 5 (September): 1859–1898. https://doi.org/10.3982/ECTA16410.

MacKinnon, James G., and Halbert White. 1985. "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29, no. 3 (September): 305–325. https://doi.org/10.1016/0304-4076(85)90158-7.

Müller, Ulrich K. 2023. "A More Robust t-Test." Forthcoming, February. https://doi.org/10.1162/rest_a_01291.

Satterthwaite, F. E. 1946. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2, no. 6 (December): 110–114. https://doi.org/10.2307/3002019.