

Nonstandard Standard Errors

One Pager

Siddhartha Basu

March 27, 2017

The Bias of Robust Standard Errors

Theory

We know from before that $\hat{\beta}$, the OLS estimate of β has the following distribution:

$$\sqrt{N}(\hat{\beta} - \beta) \sim \mathcal{N}(0, \Omega)$$

In the case of homoskedasticity the conventional estimator that we have for Ω is:

$$\hat{\Omega}_c = (X'X)^{-1} \hat{\sigma}^2 = (X'X)^{-1} \sum \frac{\epsilon_i^2}{N}$$

In the case of heteroskedasticity the robust estimator that we have for Ω is:

$$\hat{\Omega}_r = (X'X)^{-1} \left(\sum \frac{X_i X_i' \epsilon_i^2}{N} \right) (X'X)^{-1}$$

It is well known that $\hat{\Omega}_c$ is biased. This can be fixed with a $N - K$ correction for the degrees of freedom. It is less well known that in cases of mild heteroskedasticity, robust standard errors are actually more biased than conventional standard errors. This can be shown theoretically for simple cases such as bivariate regression using the hat matrix ($H = X(X'X)^{-1}X'$, $h_{ii} = X_i'(X'X)^{-1}X_i$). More generally, simulation and Monte-Carlo based methods show this.

The main theorized fixes to the situation are to get better estimators of $\psi_i = \mathbb{E}[\epsilon\epsilon']_{ii}$. The estimator $\hat{\Omega}_r$ sets $\hat{\psi}_i = \hat{\epsilon}_i^2$ as proposed by White (1980). McKinnon and White (1985) propose the following:

$$HC_0 : \hat{\psi}_i = \hat{\epsilon}_i^2$$

$$HC_1 : \hat{\psi}_i = \frac{N}{N - K} \hat{\epsilon}_i^2$$

$$HC_2 : \hat{\psi}_i = \frac{N}{1 - h_{ii}} \hat{\epsilon}_i^2$$

$$HC_3 : \hat{\psi}_i = \frac{N}{(1 - h_{ii})^2} \hat{\epsilon}_i^2$$

Bootstrap

Bootstrap: sample N observations from our data with replacement, and estimate $\hat{\beta}$. The standard deviation of $\hat{\beta}$ across these draws is the bootstrap SE. There are a few flavors of the bootstrap:

- The "pairs" or "nonparametric bootstrap": draw pairs of $\{y_i, X_i\}$.

- The "parametric bootstrap": keep X_i fixed, and draw residuals $\hat{\epsilon}_i$, to make synthetic y_i . Mimics non-stochastic regressors and ensures that X_i and residuals are independent.
- "Wild bootstrap": Draw $X_i'\hat{\beta} + \hat{\epsilon}_i$ half the time, and $X_i'\hat{\beta} - \hat{\epsilon}_i$ the other half. Preserves relationship b/w residual variances and X_i observed in the sample

The bootstrap is good for finding the asymptotic distribution when it is hard to compute. When a test statistic is asymptotically pivotal (asymptotic distribution does not depend on any unknown parameters), the bootstrap provides **asymptotic refinement** (the bootstrap distribution is closer to the finite sample distribution of interest than the asymptotic approximation). The t-statistic is asymptotically pivotal. Standard errors are not, however we usually bootstrap standard errors and compute a t-statistic off them.

In the t statistic case, reject hypothesis when original test statistic higher than 95th percentile of absolute values from the bootstrap distribution. As applied researchers, we don't like bootstrapping pivotal statistics, since we are not interested in formal hypothesis testing. Focus on bias corrections like HC_1 to HC_3 .

Simulation Example

To see how robust and conventional standard error estimators vary, we simulate data of the form:

$$y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

Where D_i is a dummy variable. Heteroskedasticity in this concept means that the variances in the $D_i = 1$ and $D_i = 0$ populations are different. In this case, the White standard error for β_1 is the same as the Welch t-statistic denominator: $\frac{S_0^2}{N_0} + \frac{S_1^2}{N_1}$

On the other hand, the conventional estimator pools subsamples, which is efficient when the two variances are the same. When treatment and control are balanced (1/2 in each), all estimators (conventional, White, $HC_{1,2,3}$) differ little.

Let's take an example with $N = 30$, $N_0 = 27$ and $N_1 = 3$. We draw $e_i \sim N(0, \sigma^2)$ when $D_i = 0$, and $e_i \sim N(0, 1)$ when $D_i = 1$. The three cases we consider are high heteroskedasticity, with $\sigma = 0.5$, little heteroskedasticity with $\sigma = 0.85$, and homoskedasticity with $\sigma = 1$. We run the experiment with no true effect, and replicate it 25,000 times.

The main takeaways are:

- Lots of het, conventional standard errors only 1/2 as large as they should be. Robust SE's are bigger but still too small.
- Standard errors are themselves estimates and have a lot of sampling variability, which gets worse for robust and $HC_{1,2,3}$.
- Little het, White SE's have smaller mean than conventional, and higher sampling variance. This leads to higher rejection rates.

Generally, take results where robust standard errors fall below conventional ones with a grain of salt. In this spirit, you can also take the max of the conventional and robust standard errors as the best measure of precision.

Notation: X_i' is the i -th observation/row of X , the data matrix.

Fun facts about matrix multiplication