Nonstandard Standard Errors
One Pager

Siddhartha Basu

March 20, 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

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.

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