Heteroscedasticity-consistent standard errors

From Wikipedia, the free encyclopedia

The topic of **heteroscedasticity-consistent (HC) standard errors** arises in statistics and econometrics in the context of linear regression as well as time series analysis. These are also known as **Eicker–Huber–White standard errors** (also **Huber–White standard errors** or **White standard errors**),[1] to recognize the contributions of Friedhelm Eicker,[2] Peter J. Huber,[3] and Halbert White.[4]

In regression and time-series modelling, basic forms of models make use of the assumption that the errors or disturbances u_i have the same variance across all observation points. When this is not the case, the errors are said to be heteroscedastic, or to have heteroscedasticity, and this behaviour will be reflected in the residuals $\widehat{u_i}$ estimated from a fitted model. Heteroscedasticity-consistent standard errors are used to allow the fitting of a model that does contain heteroscedastic residuals. The first such approach was proposed by Huber (1967), and further improved procedures have been produced since for cross-sectional data, time-series data and GARCH estimation.

Contents

Definition

Assume that we are studying the linear regression model

$$Y = X'\beta + U,$$

where X is the vector of explanatory variables and β is a $k \times 1$ column vector of parameters to be estimated.

The ordinary least squares (OLS) estimator is

$$\widehat{\beta}_{OLS} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}.$$

where $\mathbb{X}$ denotes the matrix of stacked X_i' values observed in the data.

If the sample errors have equal variance σ^2 and are uncorrelated, then the least-squares estimate of β is BLUE (best linear unbiased estimator), and its variance is easily estimated with

$$v_{OLS}[\hat{\beta}_{OLS}] = s^2(\mathbb{X}'\mathbb{X})^{-1}, s^2 = \frac{\sum_i \hat{u}_i^2}{n-k}$$

where $\hat{u}_i$ are regression residuals.

When the assumptions of $E[uu'] = \sigma^2 I_n$ are violated, the OLS estimator loses its desirable properties. Indeed,

$$V[\hat{\beta}_{OLS}] = V[(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}] = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\Sigma\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}$$

where $\Sigma = V[u]$.

While the OLS point estimator remains unbiased, it is not "best" in the sense of having minimum mean square error, and the OLS variance estimator $v_{OLS}[\hat{\beta}_{OLS}]$ does not provide a consistent estimate of the variance of the OLS estimates.

For any non-linear model (for instance Logit and Probit models), however, heteroscedasticity has more severe consequences: the maximum likelihood estimates of the parameters will be biased (in an unknown direction), as well as inconsistent (unless the likelihood function is modified to correctly take into account the precise form of heteroscedasticity).[5] As pointed out by Greene, "simply computing a robust covariance matrix for an otherwise inconsistent estimator does not give it redemption."[6]

Eicker's heteroscedasticity-consistent estimator

If the regression errors u_i are independent, but have distinct variances σ_i^2, then $\Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ which can be estimated with $\hat{\sigma}_i^2 = \hat{u}_i^2$. This provides White's (1980) estimator, often referred to as *HCE* (heteroscedasticity-consistent estimator):

$$\begin{aligned} v_{HCE}[\hat{\beta}_{OLS}] &= \frac{1}{n}\left(\frac{1}{n}\sum_i X_i X_i'\right)^{-1}\left(\frac{1}{n}\sum_i X_i X_i' \hat{u}_i^2\right)\left(\frac{1}{n}\sum_i X_i X_i'\right)^{-1} \\ &= (\mathbb{X}'\mathbb{X})^{-1}(\mathbb{X}' \operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)\mathbb{X})(\mathbb{X}'\mathbb{X})^{-1}, \end{aligned}$$

where as above $\mathbb{X}$ denotes the matrix of stacked X_i' values from the data. The estimator can be derived in terms of the generalized method of moments (GMM).

Note that also often discussed in the literature (including in White's paper itself) is the covariance matrix $\hat{\Omega}_n$ of the $\sqrt{n}$-consistent limiting distribution:

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Omega),$$

where,

$$\Omega = E[XX']^{-1} Var[Xu] E[XX']^{-1},$$

and

$$\begin{aligned} \hat{\Omega}_n &= \left(\frac{1}{n}\sum_i X_i X_i'\right)^{-1}\left(\frac{1}{n}\sum_i X_i X_i' \hat{u}_i^2\right)\left(\frac{1}{n}\sum_i X_i X_i'\right)^{-1} \\ &= n(\mathbb{X}'\mathbb{X})^{-1}(\mathbb{X}' \operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)\mathbb{X})(\mathbb{X}'\mathbb{X})^{-1}. \end{aligned}$$

Thus,

$$\hat{\Omega}_n = n \cdot v_{HCE}[\hat{\beta}_{OLS}]$$

and

$$\widehat{Var}[Xu] = \frac{1}{n}\sum_i X_i X_i' \hat{u}_i^2 = \frac{1}{n}\mathbb{X}' \operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)\mathbb{X}\,.$$

Precisely which covariance matrix is of concern should be a matter of context.

Alternative estimators have been proposed in MacKinnon & White (1985) that correct for unequal variances of regression residuals due to different leverage. Unlike the asymptotic White's estimator, their estimators are unbiased when the data are homoscedastic.

See also

- Generalized least squares
- Generalized estimating equations
- White test — a test for whether heteroscedasticity is present.

Software

- EViews: EViews version 8 offers three different methods for robust least squares: M-estimation (Huber, 1973), S-estimation (Rousseeuw and Yohai, 1984), and MM-estimation (Yohai 1987).[7]
- R: the `sandwich` package via the `vcovHC()` command.[8][9]
- RATS: `robusterrors` option is available in many of the regression and optimization commands (`linreg`, `nlls`, etc.).
- Stata: `robust` option applicable in many pseudo-likelihood based procedures.[10]

References

1. Kleiber, C.; Zeileis, A. (2006). "Applied Econometrics with R" (PDF). *UseR-2006 conference*. Archived from the original (PDF) on April 22, 2007.
2. Eicker, Friedhelm (1967). "Limit Theorems for Regression with Unequal and Dependent Errors". *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. pp. 59–82. MR 0214223. Zbl 0217.51201.
3. Huber, Peter J. (1967). "The behavior of maximum likelihood estimates under nonstandard conditions". *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. pp. 221–233. MR 0216620. Zbl 0212.21504.
4. White, Halbert (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". *Econometrica*. **48** (4): 817–838. doi:10.2307/1912934. JSTOR 1912934. MR 575027.
5. Giles, Dave (May 8, 2013). "Robust Standard Errors for Nonlinear Models". *Econometrics Beat*.
6. Greene, William H. (2012). *Econometric Analysis* (Seventh ed.). Boston: Pearson Education. pp. 692–693. ISBN 978-0-273-75356-8.
7. http://www.eviews.com/EViews8/ev8ecrobust_n.html
8. sandwich: Robust Covariance Matrix Estimators (https://cran.r-project.org/web/packages/sandwich/index.html)
9. Kleiber, Christian; Zeileis, Achim (2008). *Applied Econometrics with R*. New York: Springer. pp. 106–110. ISBN 978-0-387-77316-2.
10. See online help for `_robust` (https://www.stata.com/manuals13/p_robust.pdf) option and `regress` (https://www.stata.com/manuals13/rregress.pdf) command.

- Hayes, Andrew F.; Cai, Li (2007). "Using heteroscedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation". *Behavior Research Methods*. **39** (4): 709–722. doi:10.3758/BF03192961.
- MacKinnon, James G.; White, Halbert (1985). "Some Heteroskedastic-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties". *Journal of Econometrics*. **29** (29): 305–325.

doi:10.1016/0304-4076(85)90158-7.
- Greene, William (1998). *Econometric Analysis*. Prentice Hall.

Categories: Regression analysis | Simultaneous equation methods (econometrics)
