Intermediate Econometrics

Professor: Xiaojun Song

Timekeeper: Rui Zhou

Spring 2023

目录

7 Heteroskedasticity **2**

7.1 Properties of OLS under Heteroskedasticity 2

7.2 Testing for Heteroskedasticity 4

7.3 Weighted Least Squares Estimation 5

7 Heteroskedasticity

7.1 Properties of OLS under Heteroskedasticity

Assumptions for **Homoskedasticity**: $Var(u_i|x_i) = \sigma^2$.

7.1.1 Unbiasedness & Consistency

If we require OLS estimators to have some "good" properties,

- Unbiasedness: $E(u_i|x_i) = 0$ (zero conditional mean)
- Consistency: $E(u_i|x_i) = 0$ (zero conditional mean)

Based on those, even with heteroskedasticity,

- OLS still unbiased and consistent under heteroskedasticity;
- Interpretation of R^2 is not changed.

$$\begin{aligned}
\text{plim}_{\,n\to\infty} \frac{1}{n} SSR &= \sigma_u^2 \\
\text{plim}_{\,n\to\infty} \frac{1}{n} SST &= \sigma_y^2 \\
R^2 = 1 - \frac{SSR}{SST} &\xrightarrow{p} 1 - \frac{\sigma_u^2}{\sigma_y^2}
\end{aligned}$$

This is because SSR and SST measure the unconditional variance of error and DV respectively. Unconditional error variance is **unaffected** by heteroskedasticity (which refers to the conditional error variance). However, heteroskedasticity invalidates variance formulas for OLS estimators, and OLS is no longer the Best Linear Unbiased Estimator (BLUE).

7.1.2 Variance

Consider the simple linear regression $y_i = \beta_0 + \beta_1 x_i + u_i$, with $Var(u_i|x_i) = \sigma_i^2$ representing the general form for heteroskedasticity.

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
Var(\hat{\beta}_1|\vec{x}) &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var(u_i|x_i)}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}
\end{aligned}$$

Note that $Var(\hat{\beta}_1|\vec{x}) = \frac{\sum_{i=1}^n (x_i-\bar{x})^2\sigma_i^2}{(\sum_{i=1}^n (x_i-\bar{x})^2)^2}$ is valid regardless of heteroskedasticity or homoskedasticity, while $Var(\hat{\beta}_1|\vec{x}) = \frac{\sigma_i^2}{\sum_{i=1}^n (x_i-\bar{x})^2}$ holds only for homoskedasticity.

However, from a given sample, we can only make inference based on unbiased estimators for σ_i^2.

$$\widehat{Var(\hat{\beta}_1)} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \hat{u}_i^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}$$
$$se(\hat{\beta}_1) = \sqrt{\widehat{Var(\hat{\beta}_1)}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \hat{u}_i^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}}$$

Note that this version of $se(\hat{\beta}_1)$ here is heteroskedasticity-robust! And all formulas are only valid in large samples.

$$\text{plim}_{n\to\infty}\ nVar(\hat{\beta}_1|\vec{x}) = \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{\left(\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{E\left((x-\mu_x)^2\sigma_i^2\right)}{(\sigma_x^2)^2}$$
$$\text{plim}_{n\to\infty}\ n\widehat{Var(\hat{\beta}_1|\vec{x})} = \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \hat{u}_i^2}{\left(\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{E\left((x-\mu_x)^2\sigma_i^2\right)}{(\sigma_x^2)^2}$$
$$\Longrightarrow \quad \text{plim}_{n\to\infty}\ n\widehat{Var(\hat{\beta}_1|\vec{x})} = \text{plim}_{n\to\infty}\ nVar(\hat{\beta}_1|\vec{x})$$

The first two equations above can be proved via LLN and CLT.

$$\text{plim}_{n\to\infty}\ \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2\sigma_i^2 = \text{plim}_{n\to\infty}\ \frac{1}{n}\sum_{i=1}^n (x_i - \mu_x)^2\sigma_i^2 \overset{LLN}{\Longrightarrow} E\left((x-\mu_x)^2\sigma_i^2\right)$$
$$\text{plim}_{n\to\infty}\ \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2\hat{u}_i^2 = \text{plim}_{n\to\infty}\ \frac{1}{n}\sum_{i=1}^n (x_i - \mu_x)^2\sigma_i^2 \overset{LLN}{\Longrightarrow} E\left((x-\mu_x)^2\sigma_i^2\right)$$
$$\begin{aligned} E\left((x-\mu_x)^2\sigma_i^2\right) =& E\left[E\left[(x-\mu_x)^2\sigma_i^2|x_i\right]\right] \\ =& E\left[(x-\mu_x)^2 E(\sigma_i^2|x_i)\right] \\ =& E\left[(x-\mu_x)^2\sigma_i^2\right] \end{aligned}$$

In the MLR case,

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i \widehat{Var(\hat{\beta}_j|\vec{x})} = \frac{\sum_{i=1}^n \hat{r}_{ij}\hat{u}_i^2}{SSR_j^2}$$

Its corresponding $se(\hat{\beta}_j|\vec{x})$ is called White/Huber/Elcker standard errors. They involve the squared residuals from the regression and from a regression of x_j on all other exlanatory variables.

7.1.3 Hypothesis Test

Under the heteroskedasticity-robust OLS standard error, the usual t test is valid *asymptotically* (If with heteroskedasticity, the t statistic constructed on usual standard error is no longer valid.). The heteroskedasticity-robust t statistic differs with the traditional one only in its version of standard error!

$$t = \frac{\text{estimate - hypothesized value}}{\text{standard error}}$$
$$t_{\text{robust}} = \frac{\text{estimate - hypothesized value}}{\text{heteroskedasticity-robust standard error}}$$

Generally speaking (NOT theoretically), heteroskedasticity-robust standard errors may be *slightly larger* than their non-robust counterparts. If they show great difference, strong heteroskedasticity is indicated.

However, the usual F statistic does not work under heteroskedasticity, but heteroskedasticity-robust versions are available in statistical softwares (Too complicated to master).

7.2 Testing for Heteroskedasticity

7.2.1 Breusch-Pagan Test

Under MLR.4,

$$\begin{aligned} H_0 &: Var(u|\vec{x}) = \sigma^2 \\ Var(u|\vec{x}) &= E(u^2|\vec{x}) - [E(u|\vec{x})]^2 = E(u^2|\vec{x}) \\ \implies \quad & E(u^2|\vec{x}) = E(u^2) = \sigma^2 \end{aligned}$$

which means the mean of u^2 must not vary with $x_1, x_2, ..., x_k$.

Consider the following equation,

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + v$$

If the null hypothesis holds, then the joint hypothesis $\delta_1 = \delta_2 = ... = \delta_k = 0$ cannot be rejected. However, the "real" u^2 cannot be observed, so fine to substitute u^2 with $\hat{u}^2$.

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + error$$

The error term incorporates two parts, $error_i = v_i + \hat{u}_i^2 - u_i^2$. With sufficiently large samples, $\hat{u}_i^2 - u_i^2$ is thought to converge to 0.

Accordingly, conduct the F test. Construct the F statistic

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n-k-1)} \sim F_{k,n-k-1}$$

whose alternative is the Lagrange Multiplier Test with LM statistic.

$$LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_k$$

Note that a small $R^2_{\hat{u}^2}$ may not correspond to a small LM statistic, since the latter is then multiplied by the sample size n, which is potentially large.

Remarks

- Taking logarithm form of variables can alleviate the issue of heteroskedasticity to some extent, and this is accomplished mainly through the dependent variable, instead of the explanatory variable.

- BP Test only considers the linear function of $\vec{x}$.
- More on F and LM.
  - $F_{k,n-k-1} \overset{n\to\infty}{\Longrightarrow} F_{k,\infty} \overset{a}{\sim} \chi^2_k/k$
  - And if $X_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, then $\frac{1}{n}\sum_{i=1}^n X_i^2 \sim \chi^2_n/n$. When n approaches infinity, $\frac{1}{n}\sum_{i=1} X_i^2 \overset{p}{\to} E(X_i^2) = Var(X_i) = 1$, then $\chi^2_n/n \overset{p}{\to} 1, n \to \infty$.
  - $LM \overset{a}{\sim} k \cdot F, n \to \infty$

7.2.2 White Test

Since BP Test only considers the linear function of $\vec{x}$, White Test sees to it and take more higher-level items of independent variables into account. Regress squared residuals on all explanatory variables, their squaresm and interactions. Take a 3-explanatory-variable case for example:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 + \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + error$$

The main drawback is that, the items needed to taken into account will expand greatly with more explanatory variables. Then more estimators are needed to be estimated, and eventually the power of the test shrinks.

Alternative form of the White Test:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + error$$

Note that it is the estimated $\hat{y}$ that is incorporated into the regression. And the equation makes sense only when $\hat{y}$, instead of y itself is used.

The predicted $\hat{y}$ and its square implicitly contain all of these terms. Thus ,the regression above **indirectly** tests the dependence of the squared residuals on the explanatory variables, their squares, and interactions. And this alternative form of White Test is also called the “Special Case” for White Test.

7.3 Weighted Least Squares Estimation

Suppose that, heteroskedasticity is known up to a multiplicative constant, i.e., the functional form of the heteroskedasticity is known.

$$Var(u_i|\vec{x}_i) = \sigma^2 \cdot h(\vec{x}_i), h(\vec{x}_i) = h_i > 0$$

Assumptions of the multiplicative constant are kind of arbitraty. Discussions about underlying beliefs are needed.

Note that $h(\vec{x}_i) > 0$, which is the primary condition to check! Meanwhile, this constraint limits WLS’s application.

7.3.1 Model Transformation for WLS

Naturally, we can get

$$Var\left(\frac{u_i}{\sqrt{h_i}}\middle|\vec{x}_i\right) = \frac{1}{h_i}\cdot Var(u_i|\vec{x}_i) = \sigma^2$$

From the traditional MLR model, we can get a transformed model to tackle with heteroskedasticity, taking the funtional form into account.

$$\begin{aligned}
& y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i \\
\Longrightarrow \quad & \left[\frac{y_i}{\sqrt{h_i}}\right] = \beta_0\left[\frac{1}{\sqrt{h_i}}\right] + \beta_1\left[\frac{x_{i1}}{\sqrt{h_i}}\right] + ... + \beta_k\left[\frac{x_{ik}}{\sqrt{h_i}}\right] + \left[\frac{u_i}{\sqrt{h_i}}\right] \\
\Longleftrightarrow \quad & y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + ... + \beta_k x_{ik}^* + u_i^*
\end{aligned}$$

Note that h_i is a function of $\vec{x}_i$. In such transformed model, here is no intercept term. Also, the transformation factor does not necessarily have to be exactly $\frac{1}{\sqrt{h_i}}$, fine to be any **multiple** of $\frac{1}{\sqrt{h_i}}$.

For the transformed model, conduct regression using OLS.

$$\begin{aligned}
& \min_{b_0,b_1,...,b_k} \sum_{i=1}^{n}\left(\left[\frac{y_i}{\sqrt{h_i}}\right] - b_0\left[\frac{1}{\sqrt{h_i}}\right] - b_1\left[\frac{x_{i1}}{\sqrt{h_i}}\right] - ... - b_k\left[\frac{x_{ik}}{\sqrt{h_i}}\right]\right)^2 \\
\Longleftrightarrow \quad & \min_{b_0,b_1,...,b_k} \sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_{i1} - ... - b_k x_{ik}\right)^2 / h_i
\end{aligned}$$

If h_i is a constant, i.e., independent of $\vec{x}_i$, then this algorithm of WLS coincides with OLS. The "weighted" in WLS lies in the h_i for i. Observations with a larger variance are less informative than those with smaller variance and therefore should get less weight. WLS estimators sometimes have considerably smaller standard errors, which is line with the expectation that they are more efficient.

WLS is a special case of generalized least squares. Under WLS, the corresponding deterministic coefficient is weighted R^2, $R^2 = 1 - \frac{wSSR}{wSST}$. (Not for you to master)

Moreover, if the other Gauss-Markov assumptions hold as well, OLS applied to the transformed model is the BLUE (Best Linear Unbiased Estimator).

Lastly, set a powerful reminder to yourself that, the interpretation of coefficients should come back to the **traditional** model; our transformation only aims to improve the effectiveness of estimation.

Also, note that both heteroskedasticity-robust standard error and WLS are ways to cope with issues of heteroskedasticity and propose a correction. If the estimated slope coefficients are close to each other under two methods (both in magnitude and error), just be confident about the cross-checked validation.

7.3.2 Important Special Case for WLS

If the observations are reported as average at the grand level (such as city, country, etc.), they should be **weighted** by the **size of the unit**. Suppose for firm i with number of employees to be m_i, and the indicator $y_{i,e}$ locates individual e in firm i.

$$\begin{aligned} & y_{i,e} = \beta_0 + \beta_1 x_{1,i,e} + \beta_2 x_{2,i,e} + \beta_3 x_{3,i} + u_{i,e} \\ \iff \quad & \frac{1}{m_i} \cdot \sum_{e=1}^{m_i} y_{i,e} = \beta_0 + \beta_1 \cdot \frac{1}{m_i} \cdot \sum_{e=1}^{m_i} x_{1,i,e} + \beta_2 \cdot \frac{1}{m_i} \cdot \sum_{e=1}^{m_i} x_{2,i,e} + \beta_3 x_{3,i} + \frac{1}{m_i} \cdot \sum_{e=1}^{m_i} u_{i,e} \end{aligned}$$

Thus for an entity i,

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{1,i} + \beta_2 \bar{x}_{2,i} + \beta_3 x_{3,i} + \bar{u}_i$$

It's intriguing that here $\bar{u}_i$ correlates with the sum-up level, i.e., size of the unit m_i.

$$Var(\bar{u}_i) = Var(\frac{1}{m_i} \cdot \sum_{e=1}^{m_i} u_{i,e}) = \frac{\sigma^2}{m_i}$$

where $u_{i,e}$ is i.i.d. with variance σ^2, i.e., errors are homoskedastic at the individual-level.

In sum, if errors are homoskedastic at the individual-level, WLS with weights equal to entity size m_i should be used. If assumption of homoskedasticity at the individual-level is not exactly satisfied, one can calculate robust standard errors after WLS (i.e., for the transformed model).

7.3.3 Unknown Heteroskedasticity Function in WLS

What we've discussed before comes with a strong assumption of known $h(\vec{x}_i)$. If $h(\vec{x})$ is unknown, given that $h(\vec{x}) > 0$ must hold, we can generally assume the following functional form:

$$Var(u|\vec{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + ... + \delta_k x_k) = \sigma^2 h(\vec{x})$$

where exp-function is used to ensure positivity.

Since $E(u^2|\vec{x}) = Var(u|\vec{x})$, we can write it as:

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + ... + \delta_k x_k) = \sigma^2 h(\vec{x}) \cdot v$$

where v is the multiplicative error, assume to be independent of the explanatory variables. Then,

$$\Longrightarrow \log(u^2) = \alpha_0 + \delta_1 x_1 + ... + \delta_k x_k + e$$

Substitute u^2 with sample-version $\hat{u}^2$.

$$\begin{aligned} & \log(\hat{u}^2) = \hat{\alpha}_0 + \hat{\delta}_1 x_1 + ... + \hat{\delta}_k x_k + error \\ \Longrightarrow \quad & \hat{h}_i = \exp(\hat{\alpha}_0 + \hat{\delta}_1 x_1 + ... + \hat{\delta}_k x_k) \end{aligned}$$

And then $\hat{h}_i$ can be further used as weights in WLS. (Note that $\hat{h}_i$ may be a multiple of h_i, instead of right just the h_i.)

7.3.4 Misspecified Heteroskedasticity Function

If misspecified, WLS is still consistent under MLR.1~MLR.4. Convince yourself again that the heteroskedasticity has nothing to do with unbiasedness and consistency. However, if OLS and WLS produce very different estimates, this typically indicates that some other assumptions (e.g. MLR.4) are wrong.

If with strong heteroskedasticity, still often better to use a wrong form of heteroskedasticity in order to increase efficiency.

7.3.5 WLS in LPM

We have discussed heteroskedasticity in Linear Probability Model (LPM). Now we revisit it and see how WLS is summoned to cope with it.

$$\begin{aligned} & \Pr(y = 1|\vec{x}) = p(x) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \\ \Longrightarrow \quad & Var(y|\vec{x}) = p(x)[1 - (p(x)] \end{aligned}$$

In the LPM, the exact form of heteroskedasticity is known! Thus,

$$\Longrightarrow \hat{h}_i = \hat{y}_i(1 - \hat{y}_i)$$

Remarks

- Infeasible if LPM predictions are below zero or greater than one. Otherwise, $\hat{h}_i$ would be non-positive.
  - If such cases are rare, they may be adjusted to values such as .01/.99 to stick with WLS.
  - Otherwise, probably better to use OLS with robust standard errors.