TA session2# 10

Jun Sakamoto

December 12,2018

Contents

1 Robust standard error

Huber-White standard error

We will see that OLS estimators are unbiased and consistent in the presence of heteroskedasticity, but they are not efficient and the estimated standard errors are inconsistent, so test statistics using the standard error are not valid. Huber-White standard error is one of the most popular robust standard error.
We consider the following model.

$$
\begin{aligned}
y_i &= X_i\beta + \epsilon_i \\
E(\epsilon_i) &= 0 \\
V(\epsilon_i) &= \sigma_i^2 \\
Cov(\epsilon_i, \epsilon_j) &= 0
\end{aligned}
$$

So, the variance covariance matrix is

$$\Sigma = E[\epsilon_i \epsilon_i'] = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n^2 \end{pmatrix}$$

Above matrix is a true variance covariance matrix. However, as in many other problems, Σ is unknown. One common way to solve this problem is to estimate Σ empirically: First, estimate an OLS model, second, obtain residuals, and third, estimate Σ

$$\hat{\Sigma} = \begin{pmatrix} u_1^2 & 0 & \cdots & 0 \\ 0 & \hat{u}_2^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \hat{u}_n^2 \end{pmatrix}$$

Therefore, we can estimate the variances of OLS estimators (and standard errors) by using $\hat{\Sigma}$

$$Var(\hat{\beta}) = (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1} \tag{1}$$

Standard errors based on this procedure are called (heteroskedasticity) robust standard errors or White-Huber standard errors. Or it is also known as the sandwich estimator of variance

Cluster robust standard error

Sometimes, we may impose assumptions on the structure of the heteroskedasticity. For instance, if we suspect that the variance is homoskedastic within a group but not across groups, then we obtain residuals for all observations and calculate average residuals for each group.
Consider the following regression.

$$\begin{aligned} y_{ig} &= x_i' \beta + u_{ig} \\ u_{ig} &= v_g + e_{ig} \\ E[e_{ig}|X] &= 0 \\ V[e_{ig}|X] &= \sigma_e^2 \\ Cov[e_{ig}, e_{jg}|X] &= 0 \\ E[v_g|X] &= 0 \\ V[v_g|X] &= \sigma_v^2 \\ Cov[e_{ig}, v_g|X] &= 0 \end{aligned}$$

The resulting variance-covariance structure within each cluster g is then

$$\hat{\Sigma} = V[u_g|X_g] = \sigma_u^2\Omega_g = \sigma_u^2(\rho_u 11' + (1-\rho)I) = \sigma_u^2 \begin{pmatrix} 1 & \rho_u & \cdots & \rho_u \\ \rho_u & 1 & \cdots & \rho_u \\ \vdots & \cdots & \ddots & \vdots \\ \rho_u & \cdots & \cdots & 1 \end{pmatrix} \tag{2}$$

where $\sigma_u^2 = \sigma_g^2 + \sigma_e^2$, $\rho_u = \frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2}$.
See the empirical example.

Empirical Example

"Do girl peers improve your academic performance?" F.Hu (2015) Economic letters 137 pp54-58.

This article study gender peer effects on students academic performance in China by exploiting the random within-grade-by-school variation in the share of females in the classroom.
Hu estimates the following model across subjects and separately for boys and girls

$$Y_{icgs} = \beta_0 + \beta_1 Peer_{icgs} + \beta_2 X_{icgs} + e_{icgs} \tag{3}$$

where Y_{icgs} denotes the mid-term test scores in the three compulsory subjects (math, Chinese, and English) for student i in class c of grade g of school s. $Peer_{icgs}$ measures the proportion of female students (excluding self) in the classroom in decimals. The covariate vector X_{icgs} contains background characteristics that are important determinants of students ' academic performance, including students ' age, ethnic minority status, and agricultural hukou status, parents ' education levels, number of siblings, and household income level. Throughout the analysis, I cluster the standard errors at the school level to allow for heteroskedasticity and arbitrary serial correlation across students within each school.

Table2 is result.

Table 2
Effects of female peer composition on academic performance.

	(1)	(2)	(3)	(4)	(5)
Panel A: Females					
(a) Math	1.17	3.41	5.97	5.76	4.76
	(2.36)	(3.44)	(5.92)	(5.73)	(5.56)
Observations	4537	4537	4537	4537	4537
(b) Chinese	−5.24**	−1.29	3.10	2.97	2.30
	(2.10)	(3.09)	(4.67)	(4.59)	(4.39)
Observations	4537	4537	4537	4537	4537
(c) English	−1.89	0.92	6.93	6.67	5.76
	(2.35)	(2.78)	(4.85)	(4.67)	(4.50)
Observations	4538	4538	4538	4538	4538
Panel B: Males					
(a) Math	4.49*	10.25***	17.31***	17.38***	16.53***
	(2.28)	(3.60)	(5.40)	(5.29)	(5.21)
Observations	4774	4774	4774	4774	4774
(b) Chinese	1.65	7.43**	14.27**	14.34***	13.60**
	(2.22)	(3.65)	(5.48)	(5.38)	(5.25)
Observations	4775	4775	4775	4775	4775
(c) English	2.08	8.58**	20.41***	20.63***	19.76***
	(2.08)	(3.79)	(6.54)	(6.27)	(6.09)
Observations	4773	4773	4773	4773	4773
School fixed effects	No	Yes	–	–	–
Grade-by-school fixed effects	No	No	Yes	Yes	Yes
Individual characteristics controls	No	No	No	Yes	Yes
Household characteristics controls	No	No	No	No	Yes

Individual characteristics controls include age, ethnic minority status, and agricultural hukou status. Household characteristics controls include parents' education levels, number of siblings, and dummy for low household income level. Robust standard errors clustered at the school level are shown in parentheses.
* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

No controls in the first column, but control for school fixed effects in the second column and grade-by-school fixed effects in the other three columns. In the fourth and fifth columns, I additionally control for individual and household characteristics.
The results show that female peers have positive effects on students' academic performance, especially for boys.

2 Method of moment

Moment condition is satisfied as below

$$E[g(X;\theta)] = 0 \tag{4}$$

where $g()$ is function vector with dimensional k. $X = X_1, X_2, ..., X_n$ is sample. Then moment estimator θ is defined as the paramater which satisfies below equation.

$$\frac{1}{n}\Sigma_i g(X;\theta) = 0. \tag{5}$$

3 IV; Instrumental variable methods

Some regression model has endgenous problem. Let z correlate with explanatory variable x and orthogonal with error u. Then we can estimate consistent paramator by IV.(z is called as instrumental variable.) We consider a regression model as below.

$$y = X\beta + u \tag{6}$$
$$E[z_i x_i] = M_{zx} is regular. \tag{7}$$
$$E[z_i u_i] = 0 \tag{8}$$
$$u_i, x_i, z_i \text{has forth order moment}$$

IV estimator $\hat{\beta}_{IV}$ is obtained by method of moment.

$$E[z_i u_i] = E[z_i(y_i - x_i'\beta)] = 0 \tag{9}$$

is assumed.(moment condition) Moment estimator is

$$\frac{1}{n}\Sigma_i[z_i(y_i - x_i'\beta)] = 0 \tag{10}$$
$$\hat{\beta}_{IV} = [\Sigma_i x_i']^{-1}\Sigma_i z_i x_i \tag{11}$$
$$= (Z'X)^{-1}Z'Y \tag{12}$$

Consistency

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y \tag{13}$$
$$= (Z'X)^{-1}Z'(X\beta + u) \tag{14}$$
$$= \beta + (Z'X)^{-1}Z'u \tag{15}$$
$$= \beta + (\frac{1}{n}\Sigma_i z_i x_i')^{-1}\frac{1}{n}\Sigma_i z_i u_i \tag{16}$$

By assumption and LLN,

$$\frac{1}{n}\Sigma_i z_i x_i' \to_p E[z_i x_i] = M_{zx} < \infty \tag{17}$$
$$\frac{1}{n}\Sigma_i z_i u_i \to_p E[z_i u_i] = 0 \tag{18}$$

So we can get,

$$\hat{\beta}_{IV} \to_p \beta \tag{19}$$

Asymptotic normality

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) = \sqrt{n}(Z'X)^{-1}Z'u \tag{20}$$
$$= (\frac{1}{n}\Sigma_i z_i x_i')^{-1}\frac{1}{n}\Sigma_i z_i u_i \tag{21}$$

We use $E[z_i u_i] = 0$ and $V[z_i u_i] = E[u_i^2 z_i z_i'] = S < \infty$ from assumption. Then by CLT and normality

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \to_d M_{xx}^{-1}N(0, S) = N(0, M_{xx}^{-1}SM_{xx}^{-1}) \tag{22}$$

4 2SLS; 2 steps least square

We consider the case of multiple endogenous variables, exogenous variable and IV.

$$y = X^1\beta^1 + X^2\beta^2 + u \tag{23}$$

where X^1:matrix of endogenous variables, X^2:exogenous variable and Z:IV. l is the number of IV and k is a number of endogenous variables. We can consider 3 cases.

$l > k$:over identified
$l = k$:just identified
$l < k$:under identified

Now we consider the case of over identified. Thus if IV are more than endogenous variables, Then we use 2SLS. First we estimate a below equation.

$$X^1 = Z^1\delta^1 + X^2\delta^2 + v \tag{24}$$

We generate predict values $\hat{X}^1$ by parameters of above equation. This equation is called as reduced form. We use this values in second step regression as below.

$$y = \hat{X}^1\beta^1 + X^2\beta^2 + u \tag{25}$$

Estimator of above equation is same as IV estimator which employ $\hat{X}^1$ and X^2 as IV. Next we define below.

X:A matrix of all variables.((endogenous variables)+(exogenous variables))
Z:A matrix of all exogenous variables.

$\hat{X}$:A matrix of $(\hat{X}^1, X^2)$
Predict value $\hat{X}$ is obtained by,

$$\hat{X} = Z(Z'Z)^{-1}Z'X \tag{26}$$

The coefficients of reduced form are written as below.

$$\delta = (Z'Z)^{-1}Z'X \tag{27}$$

Thus $\hat{\beta}_{2SLS}$ is

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y && (28)\\
&= ((Z(Z'Z)^{-1}Z'X)'Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)^{-1}y && (29)\\
&= (X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y && (30)\\
&= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y && (31)
\end{aligned}$$

Next we define some assumptions to confirm asymptotic properties of 2SLS.

$$\begin{aligned}
E[z_i x_i'] &= M_{zx}\text{is regular.} && (32)\\
E[z'z] &= M_{zz}\text{is regular.} && (33)\\
E[z_i u_i] &= 0 && (34)\\
u_i, x_i, z_i&\text{has forth order moment}
\end{aligned}$$

Consistency

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y && (35)\\
&= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'(X\beta + u) && (36)\\
&= \beta + (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'u && (37)\\
&= \beta + (\frac{1}{n}\Sigma_i x_i z_i'(\frac{1}{n}\Sigma_i z_i z_i')^{-1}\frac{1}{n}\Sigma_i z_i x_i')^{-1}\frac{1}{n}\Sigma_i x_i z_i'(\frac{1}{n}\Sigma_i z_i z_i')^{-1}\frac{1}{n}\Sigma_i z_i u_i && (38)
\end{aligned}$$

By LLN and assumption,

$$\frac{1}{n}\Sigma_i x_i z_i' \to_p E[x_i z_i'] = M_{zx}' < \infty \quad (39)$$

$$\frac{1}{n}\Sigma_i z_i z_i' \to_p E[z_i z_i'] = M_{zz} < \infty \quad (40)$$

$$\frac{1}{n}\Sigma_i z_i x_i' \to_p E[z_i x_i'] = M_{zx} < \infty \quad (41)$$

$$\frac{1}{n}\Sigma_i z_i u_i \to_p E[z_i u_i] = 0 \quad (42)$$

By the continuous mapping theorem,

$$(\frac{1}{n}\Sigma_i x_i z_i'(\frac{1}{n}\Sigma_i z_i z_i')^{-1}\frac{1}{n}\Sigma_i z_i x_i')^{-1} \to_p (M_{zx}' M_{zz}^{-1} M_{zx}) < \infty \quad (43)$$

$$(\frac{1}{n}\Sigma_i z_i z_i')^{-1} \to_p E[z_i z_i']^{-1} = M_{zz}^{-1} < \infty \quad (44)$$

So, we can get,

$$\hat{\beta}_{2SLS} \to_p \beta \quad (45)$$

Asymptotic normarity

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = (\frac{1}{n}\Sigma_i x_i z_i'(\frac{1}{n}\Sigma_i z_i z_i')^{-1}\frac{1}{n}\Sigma_i z_i x_i')^{-1}\frac{1}{n}\Sigma_i x_i z_i'(\frac{1}{n}\Sigma_i z_i z_i')^{-1}\frac{1}{\sqrt{n}}\Sigma_i z_i u_i \quad (46)$$

By the LLN,

$$(\frac{1}{n}\Sigma_i x_i z_i'(\frac{1}{n}\Sigma_i z_i z_i')^{-1}\frac{1}{n}\Sigma_i z_i x_i')^{-1} \to_p (M_{zx}' M_{zz}^{-1} M_{zx})^{-1} \quad (47)$$

$$\frac{1}{n}\Sigma_i z_i x_i' \to_p M_{zx} \quad (48)$$

$$(\frac{1}{n}\Sigma_i z_i z_i')^{-1} \to_p E[z_i z_i']^{-1} = M_{zz}^{-1} \quad (49)$$

$E[z_i u_i] = 0, E[u_i^2 z_i z_i'] < \infty$ and by the CLT,

$$\frac{1}{\sqrt{n}}\Sigma_i z_i u_i \to_d N(0, E[u_i^2 z_i z_i']) = N(0, \Omega) \quad (50)$$

Thus, by the CLT,

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \rightarrow_d (M_{zx}' M_{zz}^{-1} M_{zx})^{-1} M_{zx} M_{zz}^{-1} N(0, \Omega) \tag{51}$$

$$= N(0, (M_{zx}' M_{zz}^{-1} M_{zx})^{-1} M_{zx} M_{zz}^{-1} \Omega M_{zz}^{-1} M_{zx} (M_{zx}' M_{zz}^{-1} M_{zx})^{-1}) \tag{52}$$

$$= N(0, V) \tag{53}$$