7. Standard errors

Econometrics II
Winter 2019
Osaka U

Shuhei Kitamura

Summary of the last lecture

Probit/logit model

Propensity score matching (PSM)

Outline

Heteroskedasticity

- Robust standard errors

Intracluster correlation

- Cluster-robust standard errors

Other commonly used standard errors

Heteroskedasticity

So far we have assumed that the variance of the error term, conditional on independent variables, is the same for all observations.

In some cases, the variance may vary across different segments of population.

- E.g., unobserved factors affecting savings increase with income.

In that case, we need to adjust standard errors.

- Otherwise, the OLS standard errors, and hence CIs and test statistics, are no longer valid.

To check heteroskedasticity, use the Breusch-Pagan (BP) test.

- In R, use `ncvTest` in the `car` package.

Robust standard errors

Consider a model

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \tag{1}$$

Where X_i is non-stochastic, and ε_i is independently (but not necessarily identically) distributed and $E[\varepsilon_i] = 0$.

The OLS estimator is written by

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2} && (2) \\
&= \frac{\sum_i (X_i - \bar{X})(\beta X_i + \varepsilon_i - \beta \bar{X})}{\sum_i (X_i - \bar{X})^2} && (3) \\
&= \beta + \frac{\sum_i (X_i - \bar{X})\varepsilon_i}{\sum_i (X_i - \bar{X})^2}. && (4)
\end{aligned}$$

Robust standard errors (cont.)

Thus

$$
\begin{aligned}
Var(\hat{\beta}) &= \frac{Var(\sum_i (X_i - \bar{X})\varepsilon_i)}{Var(\sum_i (X_i - \bar{X})^2)} && (5)\\
&= \frac{\sum_i (X_i - \bar{X})^2 Var(\varepsilon_i)}{\{\sum_i (X_i - \bar{X})^2\}^2} && (6)
\end{aligned}
$$

Where I used the independence assumption of ε_i ($\rightarrow$ the variance of the sum equals the sum of the variances), and the non-stochastic assumption of X_i.

Robust standard errors (cont.)

If $Var(\varepsilon_i) = \sigma^2$ (homoskedastic errors), then

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2}. \tag{7}$$

If $Var(\varepsilon_i) = \sigma_i^2$ (heteroskedastic errors), then

$$Var(\hat{\beta}) = \frac{\sum_i (X_i - \bar{X})^2 \sigma_i^2}{\{\sum_i (X_i - \bar{X})^2\}^2}. \tag{8}$$

White (1980) shows that, provided $N \to \infty$, we can use ${\hat{\varepsilon}_i}^2$ to replace σ_i^2, where $\hat{\varepsilon}_i$ is the OLS residuals from the regression of Y_i on X_i,

$$\widehat{Var}(\hat{\beta})_{robust} = \frac{\sum_i (X_i - \bar{X})^2 {\hat{\varepsilon}_i}^2}{\{\sum_i (X_i - \bar{X})^2\}^2}. \tag{9}$$

The resulting standard error is often called a robust standard error.

Intracluster correlation

We have also assumed that each observation is randomly drawn from the same population.

In practice, observations are likely to be correlated with each other within cluster (classroom, school, state, etc.).

The correlation can be either cross-sectional or temporal (or both):

- Cross-sectional: pupils in the same classroom at any given time, etc.
- Temporal: a pupil over time, etc.

If the assumption is violated, we need to adjust standard errors.

- Otherwise, the OLS standard errors, and hence CIs and test statistics, are no longer valid.

Intracluster correlation (cont.)

Consider a model

$$Y_{ig} = \beta X_{ig} + \gamma Z_g + u_{ig}, \tag{10}$$

Where g denotes a cluster (group).

Assume $u_{ig} = v_g + \eta_{ig}$. The *intraclass correlation coefficient* is

$$\rho_u = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}, \tag{11}$$

Where σ_v^2 and σ_η^2 are variances for v_g and η_{ig}, respectively.

The conventional OLS variance formula for $\hat{\beta}$ should be multiplied by a factor

$$1 + \left(\frac{Var(N_g)}{\bar{N}_g} + \bar{N}_g - 1 \right) \rho_k \rho_u, \tag{12}$$

Where ρ_k is the intraclass correlation of the kth regressor, and $\bar{N}_g$ is the average cluster size.

Intracluster correlation (cont.)

The square root of the expression is called the Moulton factor (Moulton 1986).

$$\sqrt{1+\left(\frac{Var(N_g)}{\bar{N}_g}+\bar{N}_g-1\right)\rho_k\rho_u}. \tag{13}$$

With intraclass correlation, the conventional OLS standard errors should be multiplied by this factor.

If $\rho_k = 0$ or $\rho_u = 0$, we go back to the conventional formula.

- When is this the case?

The Moulton factor is increasing in the intraclass correlations.

If all clusters have the equal size, the expression is simplied by

$$\sqrt{1+(N_g-1)\,\rho_k\rho_u}. \tag{14}$$

Cluster-robust standard errors

Consider the same model as before

$$Y_i = \alpha + \beta X_i + \varepsilon_i. \tag{15}$$

Assume that the independence assumption of ε_i no longer holds. Then

$$\frac{Var(\sum_i (X_i - \bar{X})\varepsilon_i)}{\{\sum_i (X_i - \bar{X})^2\}^2} = \frac{Cov(\sum_i (X_i - \bar{X})\varepsilon_i, \sum_j (X_j - \bar{X})\varepsilon_j)}{\{\sum_i (X_i - \bar{X})^2\}^2} \tag{16}$$

$$= \frac{\sum_i \sum_j Cov((X_i - \bar{X})\varepsilon_i, (X_j - \bar{X})\varepsilon_j)}{\{\sum_i (X_i - \bar{X})^2\}^2} \tag{17}$$

$$= \frac{\sum_i \sum_j (X_i - \bar{X})(X_j - \bar{X}) E[\varepsilon_i \varepsilon_j]}{\{\sum_i (X_i - \bar{X})^2\}^2}, \tag{18}$$

Where I used the product formula

$$Cov(\sum_i X_i, \sum_j Y_j) = \sum_i \sum_j Cov(X_i, Y_j) \tag{19}$$

To derive (17), and $E[\varepsilon_i] = 0$ to derive (18).

Cluster-robust standard errors (cont.)

Let's assume $E[\varepsilon_j \varepsilon_j] = 0$ unless i and j are in the same cluster. Then

$$Var(\hat{\beta}) = \frac{\sum_i \sum_j (X_i - \bar{X})(X_j - \bar{X}) E[\varepsilon_i \varepsilon_j] 1[\text{i,j}]}{\{\sum_i (X_i - \bar{X})^2\}^2}, \tag{20}$$

Where $1[i,j]$ takes value one if i and j are in the same cluster, and zero otherwise.

Provided the number of clusters goes to infinity, we can use

$$\widehat{Var}(\hat{\beta})_{cluster} = \frac{\sum_i \sum_j (X_i - \bar{X})(X_j - \bar{X}) \hat{\varepsilon}_i \hat{\varepsilon}_j 1[\text{i,j}]}{\{\sum_i (X_i - \bar{X})^2\}^2}, \tag{21}$$

Where $\hat{\varepsilon}_i$, $\hat{\varepsilon}_j$ are the OLS residuals from the regression of Y_i on X_i.

- The resulting standard error is called a cluster-robust standard error (actually, this is also robust to heteroskedasticity).
- If there is one observation in each cluster, this boils down to (9).

Typically, $\widehat{Var}(\hat{\beta})_{cluster}$ is larger than $\widehat{Var}(\hat{\beta})_{robust}$.

R exercise

Let's compute various standard errors!

Launch RStudio.

Type

```
data(card.data)
card <- card.data

reg1 <- lm(lwage ~ educ + exper + expersq + black + smsa
+ south, data=card)
reg2 <- lm(lwage ~ educ + exper + expersq + black + smsa
+ south + factor(region), data=card)

reg3 <- lm(lwage ~ educ + exper + expersq + black + smsa
+ south + factor(region), data=card)
cov <- vcovHC(reg3, type="HC1")
robust.se <- sqrt(diag(cov))
```

R exercise (cont.)

Type

```
reg4 <- felm(lwage ~ educ + exper + expersq + black +
smsa + south | region | 0 | region, data=card)

stargazer(reg1, reg2, reg3, reg4, se=list(NULL, NULL,
robust.se, NULL), type="text")
```

Comparing the conventional standard errors, robust standard errors, and cluster-robust standard errors, what do you find?

Haiku by Keisuke Hirano

T-stat looks too good
Try clustered standard errors—
Significance gone

R exercise (cont.)

Let's compute the Moulton factor.

First, compute intraclass correlations.

- Let's focus on `south` below.

Type

```
m1 <- lm(lwage ~ educ + exper + expersq + black + smsa +
south, data=card)
coef <- m1$coefficients
u <- residuals(m1)
res <- summary(m1)$coef

rho.u <- ICCest(region, u, data=card)
rho.south <- ICCest(region, south, data=card)
```

R exercise (cont.)

Type

```
by.region <- group_by(card, region)
groups <- summarize(by.region, count=n())

var.g <- var(groups[[2]])
mean.g <- mean(groups[[2]])
```

R exercise (cont.)

Finally, compute the Moulton factor and an adjusted standard error. Type

```
m.south <- sqrt(1+(var.g/mean.g + mean.g - 1) *
rho.south[[1]] * rho.u[[1]])

coef[[7]] * m.south

m2 <- felm(lwage ~ educ + exper + expersq + black + smsa
+ south | 0 | 0 | region, data=card)

stargazer(m1, m2, type="text")
```

The number of clusters

To apply the cluster-robust standard errors, the number of clusters should be large (say, > 30).

- E.g., $\#$ US states $= 50$, $\#$ Japanese prefectures $= 47$

If the cluster size is smaller than that, test statistics can be wrong.

In that case, one can apply the wild cluster bootstrap (Cameron et al. 2008) or other methods.

- `cluster.wild.glm` in the `clusterSEs` package does this for you.

R exercise (cont.)

Type

```
reg5 <- glm(lwage ~ educ + exper + expersq + black +
smsa + south, data=card)

stargazer(reg5, type="text", report=('vc*p'))

cluster.wild.glm(mod=reg5, dat=card, cluster= ~ region,
ci.level = 0.95, boot.reps = 100)
```

Discussion

When should we use a clustered standard error?

- If there is a good reason to believe that observations are correlated within cluster (e.g., individuals in a treated village).

Which cluster level should we use?

- Case-by-case. (E.g., estimating the impact of a state-level policy $\rightarrow$ cluster at the state level if there are observations within a state).
- If your data are panel, it's better to cluster at a higher level (the state level, rather than the state-by-year level).
- You need a sufficient number of clusters (> 30).

Other commonly used standard errors

Multiway clustering

- Clustering by more than one variable (e.g., state and year)

Conley standard errors (Conley 1999)

- Standard errors account for spatial autocorrelation

Summary

Heteroskedasticity

- Robust standard errors

Intracluster correlation

- Cluster-robust standard errors

Other commonly used standard errors